

21177 SEARCH REQUEST FORM

Examiner # (Mandatory): 69507 Requester's Full Name: John YL

Art Unit 1646 Location (Bldg/Room#): CM 10815 Phone (circle 305 306 (308) 4008

Serial Number: 09/371333 Results Format Preferred (circle): PAPER DISK E-MAIL

Title of Invention _____

Inventors (please provide full names): _____

Earliest Priority Date: _____

Keywords (include any known synonyms registry numbers, explanation of initialisms):

RECEIVED
OCT 26 1999
TECH/CHIEF OF STIC

Search Topic:

Please write detailed statement of the search topic, and the concept of the invention. Describe as specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples of relevant citations, authors, etc., if known. You may include a copy of the abstract and the broadcast or most relevant claim(s).

Please search SFO 712 10: 1, 2
and 7 of 09/371333.

FOR OFFICIAL USE ONLY

STAFF USE ONLY

Searcher: HART Type of Search Vendors (include cost where applicable)

Searcher Phone #: 305-9203	<input checked="" type="checkbox"/> N.A. Sequence	<input type="checkbox"/> STN
Searcher Location: 12014	<input checked="" type="checkbox"/> A.A. Sequence	<input type="checkbox"/> Questel/Orbit
Date Picked Up: 10/28/99	<input type="checkbox"/> Structure (#)	<input type="checkbox"/> Lexis/Nexis
Date Completed: 11/1/99	<input type="checkbox"/> Bibliographic	<input type="checkbox"/> WWW/Internet
Clerical Prep Time: 5	<input type="checkbox"/> Litigation1	<input checked="" type="checkbox"/> In-house sequence systems (list)
Terminal Time: 5	<input type="checkbox"/> Fulltext	<input type="checkbox"/> Dialog
Number of Databases: 11	<input type="checkbox"/> Procurement	<input type="checkbox"/> Dr. Link
	<input type="checkbox"/> Other	<input type="checkbox"/> Westlaw
		<input type="checkbox"/> Other (specify)

COMPUGEN

BEST AVAILABLE COPY

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 1999, 00:28:26 ; Search time 12.22 Seconds
(without alignments)
1262.292 Million cell updates/sec

Title: US-09-371-333-2
Perfect score: 1991
Sequence: 1 MWGRLLWPLVLGSLSGGT.....SKASAEGRGGMTHSSLLQ 385

Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database : PIR_60:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	530	26.6	399	2 148705	proteinase activat
2	509	25.6	397	2 S66518	proteinase-activat
3	508	25.5	369	2 G02131	proteinase-activat
4	503	25.3	397	2 S64709	proteinase-activat
5	488	24.5	420	2 I51667	thrombin receptor
6	473	23.8	425	2 A37912	thrombin receptor
7	456.5	22.9	432	2 A43448	thrombin receptor
8	452	22.7	427	2 S17148	alpha-thrombin rec
9	372.5	18.7	308	2 I50241	G protein-coupled
10	371	18.6	326	2 G02514	P2 purinoceptor -
11	371	18.6	328	2 JC4800	P26 receptor - hu
12	367.5	18.5	328	2 I55450	a novel G protein-
13	365	18.3	350	1 A37963	complement C5a ana
14	345.5	17.4	370	2 JC5549	heptahelical P2Y5-
15	343	17.2	380	2 I38435	angiotensin recept
16	337.5	17.0	361	2 B45680	G protein-coupled
17	332.5	16.7	388	2 JN0605	somatostatin recep
18	332.5	16.7	355	2 JQ1231	somatostatin recep
19	326.5	16.4	350	2 A39445	interleukin-8 rece
20	326.5	16.4	353	2 JC2492	interleukin-8 rece
21	325	16.3	365	2 S68679	G protein-coupled
22	325	16.3	351	2 B42009	FMRF-related recep
23	323.5	16.2	418	2 A46226	somatostatin recep
24	322	16.2	364	2 JN0763	somatostatin recep
25	321.5	16.1	359	2 JC5277	G protein-coupled
26	320.5	16.1	360	2 A53611	interleukin-8 rece
27	320.5	16.1	363	2 I57955	somatostatin recep
28	320	16.1	373	2 JC4162	P2Y receptor - bov
29	320	16.1	363	2 I57940	somatostatin recep
30	320	16.1	387	3 JC5949	galanin receptor 2
31	319.5	16.0	428	2 A44021	somatostatin recep
32	318.5	16.0	351	2 A46525	complement C5a ana
33	317	15.9	355	2 G02436	C-C chemokine rece
34	317	15.9	355	2 A57237	chemokine (C-C) re
35	316.5	15.9	384	2 A47249	brain-specific som
36	316	15.9	373	2 JC4737	P2Y1 receptor G-pr
37	315	15.8	359	2 A42656	angiotensin II rec
38	314.5	15.8	355	2 JC5067	G protein-coupled
39	314.5	15.8	428	2 S30508	probable G protein

40 312.5 15.7 359 2 A48857 AT1 angiotensin II
41 312 15.7 362 2 S33733 G protein-coupled
42 311.5 15.6 359 2 S15403 angiotensin II rec
43 310 15.6 359 2 A48921 interleukin-8 rece
44 309.5 15.5 352 2 S27357 complement C5a ana
45 309.5 15.5 359 2 JH0621 angiotensin recept

ALIGNMENTS

RESULT 1

I48705
proteinase activated receptor 2 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 02-Jul-1996
C:Accession: I48705
R:Nystedt, S.; Larsson, A.K.; Aberg, H.; Sundelin, J.
J. Biol. Chem. 270, 5950-5955, 1995
A:Title: The mouse proteinase-activated receptor-2 cDNA and gene. Molecular cloning a
A:Reference number: I48705; MUID:95197620
A:Accession: I48705
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-399 <RES>
A:Cross-references: EMBL:Z48043; NID:g663020; PID:g663021

Query Match 26.6% Score 530; DB 2; Length 399;
Best Local Similarity 34.1% Pred. No. 3.7e-34;
Matches 131; Conservative 68; Mismatches 141; Indels 44; Gaps 11;

QY 13 GFSLSGGTQT-PSYDSESGTGGDDSTPSILPAPRGYPGQVCANDSDTLELPDSSRAL 71
DB 37 GRSLIGRLTQPTI-----TGKG-----VPVEPGF-----SIDEFSASIL 71
QY 72 LGWTPTRLPALYGLVGLPANGALWV-LATQAPRLPSTMLLMNLTADILLALP 130
DB 72 TGRITTVFLPVYIIVFVIGLPSNGMALWIFLFTKKKHPAVIYMANLADLLSVWF 131
QY 131 PRIAYHGRQWPPEGAACRLATAALYGHMYSVLLAAVSLDRYALVHPL---RARAL 187
DB 132 LKISYHLGNWNYGEALCKVLIGFFYGNMYCSILFTCLSVQRYWVYVPMGPRKA- 191
QY 188 RGRRLALGLCMAAWMAAALPLTLQRTQRLARSDRVLCCHDALPLDAQASHWQPAFTC 247
DB 191 ---NIAVGSVLAIWLLIFLVTIPLYVMKQTIYPALNITTTCHDVLPEVLVGMFNFLS 247
QY 248 LALLGCFL-PLIAMLICYGATLHTLAASG-----RRYGHALRLTAVVLASAVAFFVPSN 300
DB 248 LA-IGVFLFPALLTASAYVLMIKTLRSSAMDEHSEKKRQRAIRLIITVLAIFYCFAPSN 306
QY 301 LLLLHSDPSAWGNLYGAYVPSLALSTLNSCVDPFIYVYVSAEFRDKVRAGLFORS- 360
DB 307 LLVHVFLTKTQRQSHVYALYVALCLSLNSCIDPFYVYFVSKDFRDHARNALLCRSV 366
QY 360 -PGDTVASKASAEGRGGMTHSS 382
DB 367 RTVNRMQISLSSNFKSRKSGSYSS 390

RESULT 2

S66518
proteinase-activated receptor 2 precursor - human
C:Species: Homo sapiens (man)
C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
C:Accession: S66518
R:Nystedt, S.; Emilsson, K.; Larsson, A.K.; Stroembeck, B.; Sundelin, J.
Eur. J. Biochem. 232, 84-89, 1995
A:Title: Molecular cloning and functional expression of the gene encoding the human p
A:Reference number: S66518; MUID:96048032
A:Accession: S66518
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-397 <NYS>
A:Cross-references: EMBL:249993; NID:g1008084; PID:g1008085
C:Genetics:
A:Map position: 5q13
A:Introns: 28/1
F:1-36/Domain: activation peptide #status predicted <APT>
F:37-397/Product: proteinase-activated receptor 2 #status predicted <MAT>

```
Query Match      25.6%; Score 509; DB 2; Length 397;
Best Local Similarity 33.5%; Pred. No. 1.6e-32;
Matches 120; Conservative 64; Mismatches 156; Indels 18; Gaps 7;

QY 10 LVLGFSLSGGTQPSVYDESGTGGDDSTPSILPAPRGVPGQVCANDSDTLELPDSRA 69
DB 16 LAASLSCSTGIQGTNRSSKGRSLIGKVDGTSV--TGKGVTVET-----VFSVDEFSAS 67

QY 70 LLGWYPTRLVPALYGLVLYVGLPANGALWV-LATQAPRLPSTMLMNLATADLLALA 128
DB 68 VLTGKLTTFVLPVIVTVFVVGLPNGMALWVFLRTKKHPAVIYMANLADLLSVIW 127

QY 129 LPRIAYHLRGQRPGEACRLATAALYGHMYSVLLAAVSLDRYLALVHLPLRAR 188
DB 128 FPLKIAYHIGNNWIYGEALCNVLGFFYGNMYCSILFMTCLSVQRYVWVIVNPM-GHSRK 186

QY 189 GRRALGLCMAAWLMAAALALPLTLQRTFLARSDRVLCCHDALPLDAQASHWQPAFTCL 248
DB 187 KANIAIGISLAIWLLILLVITPLVYVVKQTFIPALNITTCDDVLPQQLVGDGMFNFLSL 246

QY 249 ALGCGFL-PLLAMLLCYGATLHTLAASG-----RRYGHALRLTAVVLASAVAFFVPSNL 301
DB 247 A-IGVFLPFAFLTASAYVLMIRLSSAMDENSEKKRRAIKLIVTVLAMYLICFTPSNL 305

QY 302 LLLLHSDPSAWGNLYGAYVPSLALSTLNSCDVDFPIYVYSAEFRDKVRAGLQFORS 359
DB 306 LLVVHYFLIKSQGSHVYALYIVALCLSTLNSCIDPFYVYFVSHDFRDAKNAALLCRS 363
```

```
RESULT 3
G02131
proteinase-activated receptor 2 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
C:Accession: G02131
R:Kahn, M.L.; Coughlin, S.R.
submitted to the EMBL Data Library, September 1995
A:Reference number: H00822
A:Accession: G02131
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-369 <KAH>
A:Cross-references: EMBL:U36753; NID:g1208539; PID:g1208540
```

```
Query Match      25.5%; Score 508; DB 2; Length 369;
Best Local Similarity 36.5%; Pred. No. 1.7e-32;
Matches 110; Conservative 57; Mismatches 124; Indels 10; Gaps 5;

QY 67 SRALLGWPTRLVPALYGLVLYVGLPANGALWV-LATQAPRLPSTMLMNLATADLL 125
DB 37 SASVLGKLTTFVLPVIVTVFVVGLPNGMALWVFLRTKKHPAVIYMANLADLLS 96

QY 126 ALALPRIAYHLRGQRPGEACRLATAALYGHMYSVLLAAVSLDRYLALVHLPLRAR 185
DB 97 VIWFLKIAYHIGNNWIYGEALCNVLGFFYGNMYCSILFMTCLSVQRYVWVIVNPM-GH 155

QY 186 ALRGRALGLCMAAWLMAAALALPLTLQRTFLARSDRVLCCHDALPLDAQASHWQPAF 245
DB 156 SRKANIAIGISLAIWLLILLVITPLVYVVKQTFIPALNITTCDDVLPQQLVGDGMFNFL 215

QY 246 TCLALGCGFL-PLLAMLLCYGATLHTLAASG-----RRYGHALRLTAVVLASAVAFFV 298
```

```
DB 216 LSLA-IGVFLPFAFLTASAYVLMIRLSSAMDENSEKKRRAIKLIVTVLAMYLICFTP 274
QY 299 SNLALLLHSDPSAWGNLYGAYVPSLALSTLNSCDVDFPIYVYSAEFRDKVRAGLQF 358
DB 275 SNLALLVHYFLIKSQGSHVYALYIVALCLSTLNSCIDPFYVYFVSHDFRDAKNAALLCR 334

QY 359 S 359
DB 335 S 335

RESULT 4
S64709
proteinase-activated receptor-2 precursor - human
C:Species: Homo sapiens (man)
C>Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 13-Nov-1998
C:Accession: S64709
R:Boehm, S.K.; Kong, W.; Broemme, D.; Smeekens, S.P.; Anderson, D.C.; Connolly, A.; K
Biochem. J. 314, 1009-1016, 1996
A:Title: Molecular cloning, expression and potential functions of the human proteinase
A:Reference number: S64709
A:Accession: S64709
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-397 <BOE>
A:Cross-references: EMBL:U34038; NID:g1041728; PID:g1041729
A>Note: the authors translated the codon GTC for residue 68 as Ile and AAC for residu
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-397/Product: proteinase-activated receptor-2 #status predicted <MAT>

Query Match      25.3%; Score 503; DB 2; Length 397;
Best Local Similarity 33.2%; Pred. No. 4.6e-32;
Matches 119; Conservative 64; Mismatches 157; Indels 18; Gaps 7;

QY 10 LVLGFSLSGGTQPSVYDESGTGGDDSTPSILPAPRGVPGQVCANDSDTLELPDSRA 69
DB 16 LAASLSCSTGIQGTNRSSKGRSLIGKVDGTSV--TGKGVTVET-----VFSVDEFSAS 67

QY 70 LLGWYPTRLVPALYGLVLYVGLPANGALWV-LATQAPRLPSTMLMNLATADLLALA 128
DB 68 VLTGKLTTFVLPVIVTVFVVGLPNGMALWVFLRTKKHPAVIYMANLADLLSVIW 127

QY 129 LPRIAYHLRGQRPGEACRLATAALYGHMYSVLLAAVSLDRYLALVHLPLRAR 188
DB 128 FPLKIAYHIGNNWIYGEALCNVLGFFYGNMYCSILFMTCLSVQRYVWVIVNPM-GHSRK 186

QY 189 GRRALGLCMAAWLMAAALALPLTLQRTFLARSDRVLCCHDALPLDAQASHWQPAFTCL 248
DB 187 KANIAIGISLAIWLLILLVITPLVYVVKQTFIPALNITTCDDVLPQQLVGDGMFNFLSL 246

QY 249 ALGCGFL-PLLAMLLCYGATLHTLAASG-----RRYGHALRLTAVVLASAVAFFVPSNL 301
DB 247 A-IGVFLPFAFLTASAYVLMIRLSSAMDENSEKKRRAIKLIVTVLAMYLICFTPSNL 305

QY 302 LLLLHSDPSAWGNLYGAYVPSLALSTLNSCDVDFPIYVYSAEFRDKVRAGLQFORS 359
DB 306 LLVVHYFLIKSQGSHVYALYIVALCLSTLNSCIDPFYVYFVSHDFRDAKNAALLCRS 363

RESULT 5
I51667
thrombin receptor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Sep-1996
C:Accession: I51667
R:Gerzsten, R.E.; Chen, J.; Ishii, M.; Ishii, K.; Nanevica, T.; Turck, C.W.; Vu, T.H.
Nature 368, 648-651, 1994
A:Title: Thrombin receptor's specificity for agonist peptide is determined by its ext
A:Reference number: I51667; MUID:94195429
A:Accession: I51667
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
```


DB 410 LMPKMDTCSSHL 422

THE UNIVERSITY OF CHICAGO PRESS

THE UNIVERSITY OF CHICAGO PRESS

```
RESULT 8
SI7148
alpha-thrombin receptor - Chinese hamster
C:Species: Cricetulus griseus (Chinese hamster)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 10-Sep-1997
C:Accession: SI7148
R:Rasmussen, U.B.; Vouret-Craviari, V.; Jallat, S.; Schlesinger, Y.; Pages, G.; Paviranli
FBBS Lett. 288, 123-128, 1991
A:Title: cDNA cloning and expression of a hamster alpha-thrombin receptor coupled to Ca2+
A:Reference number: SI7148; MUID:91348247
A:Accession: SI7148
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-427 <RAG>
A:Cross-references: EMBL:X61958; NID:g940495; PID:g49538
C:Keywords: G protein-coupled receptor; transmembrane protein

Query Match 22.7%; Score 452; DB 2; Length 427;
Best Local Similarity 32.1%; Pred. No. 4.6e-28;
Matches 125; Conservative 69; Mismatches 135; Indels 60; Gaps 14;
QY 4 RLLNPLVLGFLSLG---GTQPSVYDESGST-----GGGD 36
| | | | : | | | : | | | : | | | : | | | :
DB 5 RLLN--VAAGSLGCLPILLSRPVVRQPESEMIDATVNPSSFELRNPGENTFELIPLGDDE 62
| | | | : | | | : | | | : | | | : | | | :
QY 37 DSTPSTLPAPGV-----PGQVCANDSTLELPDSSRALLGHWPTRL-VPALYGLVL 88
| | | | : | | | : | | | : | | | : | | | :
DB 63 EKNESTLPGRRAIYLNKSHSPAPLAFISE-----DASGLYSPWL--RLFIPSYITVFV 115
| | | | : | | | : | | | : | | | : | | | :
QY 89 VYGLPANGALWVLTQAP-RLPSTMLNLTADLLALPPIRIAYHLRGQRWPFGEA 147
| | | | : | | | : | | | : | | | : | | | :
DB 116 VYSLPLNIIAIAVEFLMKVKKPAVYVLMHLMADVLFVSLPLKISYFSGSDMQFGSG 175
| | | | : | | | : | | | : | | | : | | | :
QY 148 AGRLLAALYGHMYSVLLAAVSDRLYALVHLPRARALGRRLALGLCMAWLMMAAL 207
| | | | : | | | : | | | : | | | : | | | :
DB 176 MCRFATAFYCNHYSIMLTVISIDRELVAVYPIQSLSWRTLGRANETCLVIYVMAING 235
| | | | : | | | : | | | : | | | : | | | :
QY 208 ALPLTLQRTFLARSDRLVCHDALP---LDAQSHWQPAFTCLALGCLFLLAWLLCY 264
| | | | : | | | : | | | : | | | : | | | :
DB 236 VYPLLLKEQTRVPLGNTTCHDVLNETLQGFYSYFSAFSAVFLL---VPLIISTICY 292
| | | | : | | | : | | | : | | | : | | | :
QY 265 GATLHTLAAG---RRYGHALRLTAVLIASAVAFVPSNLLLLHY---SDPSPSAGN 317
| | | | : | | | : | | | : | | | : | | | :
DB 293 MSIIRCLSSSVANRSKSRALFLSAAVFCVFCVPTNVLIMHYLLLSLSD-SP-ATEK 350
| | | | : | | | : | | | : | | | : | | | :
QY 318 LYGAIVPSLALTLNSCDVPFFIYVVSAE 346
| | | | : | | | : | | | : | | | : | | | :
DB 351 AYFAYLLCVSVSSCCIDPLIYIYASSE 379
| | | | : | | | : | | | : | | | : | | | :

RESULT 9
I50241
G protein-coupled receptor 6H1 - chicken
N:Alternate names: purinoceptor 6H1
C:Species: Gallus gallus (chicken)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 04-Sep-1998
C:Accession: I50241; JC4618
R:Kaplan, M.H.; Smith, D.I.; Sundick, R.S.
J. Immunol. 151, 628-636, 1993
A:Title: Identification of a G protein coupled receptor induced in activated T cells.
A:Reference number: I50241; MUID:93329058
A:Accession: I50241
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-308 <KAP>
A:Cross-references: GB:L06109; NID:g304383; PID:g304384
R:Webb, T.E.; Kaplan, M.G.; Barnard, E.A.
Biochem. Biophys. Res. Commun. 219, 105-110, 1996
A:Title: Identification of 6H1 as a P2Y purinoceptor: P2Y5.
A:Reference number: JC4618; MUID:96190677
A:Accession: JC4618
A:Molecule type: mRNA
```

```
A:Residues: 1-308 <WEB>
A:Cross-references: DDBJ:D49712
A:Experimental source: T-cells
C:Comment: This receptor plays a role in T-cell activation.
C:Genetics:
A:Gene: P2Y5
C:Superfamily: ATP receptor P2u
C:Keywords: G protein-coupled receptor; transmembrane protein
F:15-40/Domain: transmembrane #status predicted <TM1>
F:51-74/Domain: transmembrane #status predicted <TM2>
F:89-109/Domain: transmembrane #status predicted <TM3>
F:133-153/Domain: transmembrane #status predicted <TM4>
F:177-201/Domain: transmembrane #status predicted <TM5>
F:227-248/Domain: transmembrane #status predicted <TM6>
F:269-292/Domain: transmembrane #status predicted <TM7>

Query Match 18.7%; Score 372.5; DB 2; Length 308;
Best Local Similarity 30.3%; Pred. No. 5e-22;
Matches 91; Conservative 57; Mismatches 129; Indels 23; Gaps 7;
QY 58 SDTLELPDSSRALLGHWPTRLVLPALYGLVLYVGLPANGALWVLA-TQAPRLPSTMLLM 116
| | | | : | | | : | | | : | | | : | | | :
DB 3 SSNCSTEDSFYITLYGCV-----FSWVEVLGLIANCAIYIFTFTLKVRNETTMYML 54
| | | | : | | | : | | | : | | | : | | | :
QY 117 NLATADLLALALPPIRIAYHLRGQRWPFGEAACRLATAALYGHMYSVLLAAVSLDRL 176
| | | | : | | | : | | | : | | | : | | | :
DB 55 NLATSDLLFVETLFPRI-YFVVRNWPFGDVLCISVTLFTYNNMYSILFTCISVDREL 113
| | | | : | | | : | | | : | | | : | | | :
QY 177 ALVHPLRARALGRRLALGLCMAWLMMAALALPLTLQRTFLARSDRLVCHDALPLDA 236
| | | | : | | | : | | | : | | | : | | | :
DB 114 AIVHPERSKTLTRKNARIVCAVAVITVLAGSTPASFFQSTNRQNTEQRTCFENFEST 173
| | | | : | | | : | | | : | | | : | | | :
QY 237 QASHWQPAFTCLALGCLFLLAWLLCYGATLHTL---AASGRYGHALRLTAVVLASAV 293
| | | | : | | | : | | | : | | | : | | | :
DB 174 WKTYLSRIVIFIEIVGFPIILNVTCTMVLRTLNKPLTLRSNKLKSKKVLKMFVRLV 233
| | | | : | | | : | | | : | | | : | | | :
QY 294 AF---FVPSNLLLLHYSDPSPSAGN-----LYGAIVPSLALTLNSCDVPFFIYVVS 344
| | | | : | | | : | | | : | | | : | | | :
DB 234 IFCFCFVYNTLIL-VSLMTQPTWNCVSVTAVRTWYPTLCAVSNCCDPPIYVFTS 292
| | | | : | | | : | | | : | | | : | | | :

RESULT 10
G02514
P2 purinoceptor - human
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 04-Sep-1998
C:Accession: G02514
R:Hammett, F.; Southey, M.C.; Somers, G.R.; Hutchins, A.M.; Venter, D.J.
submitted to the EMBL Data Library, March 1996
A:Reference number: H01373
A:Accession: G02514
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-326 <HAM>
A:Cross-references: EMBL:U52464; NID:g1407632; PID:g1407633
C:Genetics:
A:Gene: P2Y6
C:Superfamily: ATP receptor P2u

Query Match 18.6%; Score 371; DB 2; Length 326;
Best Local Similarity 31.4%; Pred. No. 6.9e-22;
Matches 97; Conservative 46; Mismatches 138; Indels 28; Gaps 7;
QY 65 DSSRALLGHWPTR-----LVPALYGLVLYVGLPANGALWVLTQAPRLPST-M 113
| | | | : | | | : | | | : | | | : | | | :
DB 2 DNGTGOALGLPPTTCVYRENFKQLLPVYSAVLAAGLPLNICVITQICTSRRALTRTAV 61
| | | | : | | | : | | | : | | | : | | | :
QY 114 LLMLNLTADLLALALPPIRIAYHLRGQRWPFGEAACRLATAALYGHMYSVLLAAVSLD 173
| | | | : | | | : | | | : | | | : | | | :
DB 62 YTLNLALADLLYACSLPLLIYNYAQQDHWPEGFACRLVRLFLFYANLHGSILFTCISFQ 121
| | | | : | | | : | | | : | | | : | | | :
```


A:Reference number: I52417; MUID:93192225
A:Accession: I52417
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3 <RES>
A:Cross-references: GB:S56556; GB:S56557; NID:g298577; NID:g298578
C:Genetics:
A:Gene: GDB:C5R1; C5A; C5AR
A:Cross-references: GDB:128856; OMIM:113995
A:Map position: 19q13.3-19q13.4
A:Introns: 1/3
A:Note: the list of introns may be incomplete
C:Description:
A:Function: mediates the inflammatory and chemotactic responses of polymorphonuclear neutrophils
C:Superfamily: vertebrate rhodopsin
C:Keywords: chemotaxis; G protein-coupled receptor; glycoprotein; inflammation; polymorphism
F:1-37/Domain: extracellular #status predicted <EX1>
F:38-61/Domain: transmembrane #status predicted <TM1>
F:62-71/Domain: intracellular #status predicted <IN1>
F:72-94/Domain: transmembrane #status predicted <TM2>
F:95-110/Domain: extracellular #status predicted <EX2>
F:111-132/Domain: transmembrane #status predicted <TM3>
F:133-149/Domain: intracellular #status predicted <IN2>
F:150-174/Domain: transmembrane #status predicted <TM4>
F:175-206/Domain: extracellular #status predicted <EX3>
F:207-227/Domain: transmembrane #status predicted <TM5>
F:228-242/Domain: intracellular #status predicted <IN3>
F:243-264/Domain: transmembrane #status predicted <TM6>
F:265-283/Domain: extracellular #status predicted <EX4>
F:284-307/Domain: transmembrane #status predicted <TM7>
F:308-350/Domain: intracellular #status predicted <IN4>
F:5/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 18.3%; Score 365; DB 1; Length 350;
Best Local Similarity 29.9%; Pred. No. 2.2e-21;
Matches 104; Conservative 58; Mismatches 140; Indels 46; Gaps 9;

QY 56 NDSDTLEL----PDSSRALLGWVTPRLVPALYGLVVLVGLPANGALAVLATOAPRLPS 111
DB 15 DDKDILDLNTPVDKTSNTLR---VPDILALVIFAVVFLVGLVGLNALVWVMTAFKARTIN 71

QY 112 TMLLMNLATADLLALAPRIAYHLRGQWPFGAEACRLATAALYGHMYSVLLAAVS 171
DB 72 AIWFLNLAVALDFSLALPILFTSIQVHHHPFGGAACSLPSILLNMYASILLATIS 131

QY 172 LDRLALVPLRLARLGRRLALGICMAAWLMAALALPLTLQRTFLARSD----RVL 227
DB 132 ADRFLVFPKPCQNPFRGAGLAWIACAVAW----GLALLTIPSELYRVVREYFPKVL 187

QY 228 CHDALPLDAQASHWQPAFTCLAL---LCGFLPLAMLLCYG-ATLHTLAASGRVGHAL 282
DB 188 C-----GVDSHDKRRERAVILRVLGLFLPLTLTICYTIFILLRTWSRATRTKL 241

QY 283 RLTAVVLASAVFFVESNLL-LLLHYSDSPSAWGNLYGAYVPSLALSTNSCVDPIFY 341
DB 242 KVVAVVASFFIFWLPYQVTGIMMSPLESPSTFLNKLKLSLCSVFAINCINPIIV 301

QY 342 YVSAERDKVRAGL-----FQSPGDTVASKASA 370
DB 302 VAGQFGGRRLKSLPSLLRNVLTEESVVRKSKSFTRSTVDTMAQRTQA 349

RESULT 14
JC5549
heptahelical P2Y5-like receptor - human
C:Species: Homo sapiens (man)
C:Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 17-Mar-1999
A:Accession: JC5549
R:Janssens, R.; Boeynaems, J.M.; Godart, M.; Communi, D.
Biochem. Biophys. Res. Commun. 236, 106-112, 1997
A:Title: Cloning of a human heptahelical receptor closely related to the P2Y5 receptor.

A:Reference number: JC5549; MUID:97366605
A:Accession: JC5549
A:Molecule type: DNA
A:Residues: 1-370 <JAN>
A:Cross-references: DDBJ:AF005419; NID:g2240034; PID:g2240035
C:Superfamily: ATP receptor P2u

Query Match 17.4%; Score 345.5; DB 2; Length 370;
Best Local Similarity 29.2%; Pred. No. 7.7e-20;
Matches 95; Conservative 61; Mismatches 138; Indels 31; Gaps 9;

QY 37 DSTPSILPAPRGYQGVCAQNDSTLELPDSSRALLGWVTPRLVPALYGLVVLVGLPANG 96
DB 13 DNSSSLRPR---LGNATAN--NTCIVDDSKYNLNG-----AVYSVFLIGLITS 58

QY 97 ZALWYLATQAP-RLPSTMLLMNLATADLLALAPRIAYHLRGQWPFGAEACRLATAA 155
DB 59 VSLFVFCFRMKMRSETAIFTINLAVSDLLFVCTLPFKIFYNF-NRHWPFEGDTLCKISGTA 117

QY 156 LYGHMYSVLLAAVSILDRYALVPLRLARLGRRLALGICMAAWLMAALALPLTLQRT 215
DB 118 FVTNYGSMFLTCTISVDRLAIVYPPFRSTIRTRNSAIVCAGVWILVLSGGISASLFS 177

QY 216 QTFRLARSDRVLCHDALPLDAQASHWQPAFTCLALGCLFLPLAMLLCYGATLH----- 270
DB 178 TI--NVNNAITTCFGLSKRWKYYLSKITIFIEVVGFIPLILNVSCSVVLTLRKPA 235

QY 270 TLAASGRRYGHALRTAVVLASAVFAFFVPSNLLLLHYSDSPSAWGNLY-----GAYV 323
DB 236 TLSQIGTNRKKVKLMTVHMAVFCVFPYNSVFLF-YALVRSQAITNCFLERFKIMYP 294

QY 324 PSLALSTNSCVDPIFYVVSAREP 348
DB 295 ITCLATLNLCCFDPFIYYITLESFQ 319

RESULT 15
I38435
angiotensin receptor homolog APJ - human
C:Species: Homo sapiens (man)
C:Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 15-Mar-1996
C:Accession: I38435
R:O'Dowd, B.F.; Heiber, M.; Chan, A.; Heng, H.H.; Tsui, L.; Kennedy, J.L.; Shi, X.; P
Gene 136, 355-360, 1993
A:Title: A human gene that shows identity with the gene encoding the angiotensin rece
A:Reference number: I38435; MUID:94124031
A:Accession: I38435
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-380 <RES>
A:Cross-references: EMBL:U03642; NID:g425351; PID:g425352
C:Genetics:
A:Gene: APJ
A:Map position: 11q12
A:Introns: #status absent

Query Match 17.2%; Score 343; DB 2; Length 380;
Best Local Similarity 29.5%; Pred. No. 1.2e-19;
Matches 97; Conservative 58; Mismatches 142; Indels 32; Gaps 9;

QY 79 LVPALYGLVVLVGLPANGALAWL--ATQAPRLPSTMLLMNLATADLLALAPRIAYH 136
DB 30 LIPALYMLVFLTGTTGNGLVLTWTVFRSRRKRRADFIASLAVADLTFFVTLPLWATYT 89

QY 137 LRQGWPFGEAACRLATAALYGHMYSVLLAAVSILDRYALVPLRLARLGRRLALGL 196
DB 90 YRDYDVGTFGTFCCKLSYLIFNVYASVFCGLSFDRLYLAIVRPVANARLVRSGAVA 149

QY 197 CMAAWLMAALALPLTLQRTFLARSDRVLCHDALPLDAQASH---WQPAF-TCLALLG 252
DB 150 TAVLWVLAALLAMPVWVLTGTDLNTKVCYMDYSHVATVSVSEWAVEVGLVGSYSTTVG 209

Qy 253 CFLPLAMLCYGATLHTLAASGRR-----YGHALRLTAVVVLASAVAF---FVPSNLLLL 304
Db 210 FWPFTIMLTCTCYFFTAQTIAGHFRKERIEGLRKRRLLSIIIVLVVTFALCWMPPYHLVKT 269
Qy 305 LHYSDDPSANGNLYGAVVPSL-----ALSTLNSCVDPFIYYVSAEPRDKVR----- 353
Db 270 L-YMLGSLHWPCCDFDLFLMNIFFPYCTCISYVNSCLNPFYAFDFRFRQACTSMLCGQ 328
Qy 353 ---AGLFORSFGDTVASKASAEKSGRGMG 378
Db 329 SRCAGTSHSSSGEKSASYSS--GHSQGP 355

Search completed: October 29, 1999, 02:35:48
Job time: 7642 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 1999, 11:07:23 ; Search time 11.26 Seconds
(without alignments)
966.545 Million cell updates/sec

Title: US-09-371-333-2
Perfect score: 1991
Sequence: 1 MWGRLLPLVLGFLSLGGT.....SKASAEGRSGMGTHSSLLQ 385

Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database : SwissProt_37.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	530	26.6	399	1 PAR2_MOUSE	P55086 mus musculus
2	524	26.3	397	1 PAR2_RAT	O63645 rattus norv
3	521.5	26.2	374	1 PAR3_HUMAN	O00234 homo sapien
4	509	25.6	397	1 PAR2_HUMAN	P55085 homo sapien
5	507.5	25.5	369	1 PAR3_MOUSE	O08675 mus musculus
6	488	24.5	420	1 THRR_XENLA	P47749 xenopus lae
7	480	24.1	425	1 THRR_PAPHA	P56488 papio hamad
8	473	23.8	425	1 THRR_MOUSE	P25116 homo sapien
9	468.5	23.5	430	1 THRR_MOUSE	P30558 mus musculus
10	458.5	23.0	428	1 THRR_CRILLO	O00991 cricetus
11	456.5	22.9	432	1 THRR_RAT	P26824 rattus norv
12	372.5	18.6	308	1 P2Y5_CHICK	P32250 gallus gall
13	371	18.6	328	1 P2Y6_HUMAN	Q15077 homo sapien
14	367.5	18.5	328	1 P2Y6_RAT	O63371 rattus norv
15	366	18.4	340	1 C5AR_MACMU	P79188 macaca mula
16	365	18.3	350	1 C5AR_HUMAN	P21730 homo sapien
17	362.5	18.2	340	1 C5AR_PONPY	P79234 pongo pygma
18	362	18.2	340	1 C5AR_GORGO	P79175 gorilla gor
19	362	18.2	340	1 C5AR_PANTR	P79240 pan troglod
20	354.5	17.8	339	1 GPRH_HUMAN	Q13304 homo sapien
21	346	17.4	328	1 P2Y3_CHICK	O98907 gallus gall
22	345.5	17.4	352	1 C5AR_RAT	P97520 rattus norv
23	343	17.2	380	1 APJ_HUMAN	P35414 homo sapien
24	341.5	17.2	370	1 P2Y2_HUMAN	O99677 homo sapien
25	339.5	17.1	360	1 IL8B_BOVIN	Q28003 bos taurus
26	337.5	17.0	361	1 EB12_HUMAN	P32249 homo sapien
27	332.5	16.7	355	1 IL8A_RABIT	P21109 oryctolagus
28	332.5	16.7	388	1 SSRA_HUMAN	P31391 homo sapien
29	331.5	16.6	344	1 P2Y5_HUMAN	P43657 homo sapien
30	329.5	16.5	350	1 IL8A_PANTR	P55920 pan troglod
31	328	16.5	348	1 FML1_GORGO	P79177 gorilla gor
32	328	16.5	537	1 P2Y8_XENLA	P79938 xenopus lae
33	326.5	16.4	353	1 FML1_RAT	P46090 rattus norv
34	325	16.3	351	1 FML1_HUMAN	P25090 homo sapien
35	325	16.3	365	1 P2Y4_HUMAN	P51582 homo sapien
36	323.5	16.2	350	1 IL8A_HUMAN	P25024 homo sapien
37	323.5	16.2	418	1 SSRA_HUMAN	P32745 homo sapien
38	323	16.2	357	1 G96_HUMAN	P51686 homo sapien
39	320.5	16.1	360	1 IL8B_HUMAN	P25025 homo sapien
40	320.5	16.1	363	1 SSRA_HUMAN	P35346 homo sapien
41	320	16.1	353	1 IL8B_PANTR	Q28807 pan troglod
42	320	16.1	373	1 P2YR_BOVIN	P48042 bos taurus
43	320	16.1	363	1 SSRA_RAT	P30938 rattus norv

ALIGNMENTS

RESULT 1	PAR2_MOUSE	STANDARD;	PRT;	399 AA.
ID	PAR2_MOUSE	STANDARD;	PRT;	399 AA.
AC	P55086;			
DT	01-OCT-1996 (REL. 34, CREATED)			
DT	01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)			
DE	PROTEINASE ACTIVATED RECEPTOR 2 PRECURSOR (PAR-2).			
GN	PAR2 OR GPR11 OR GPCR11.			
OS	MUS MUSCULUS (MOUSE).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;			
OC	RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 95197620.			
RA	NYSTEDT S., LARSSON A.-K., AABERG H., SUNDELIN J.;			
RT	"The mouse proteinase-activated receptor-2 cDNA and gene. Molecular cloning and functional expression."			
RL	J. BIOL. CHEM. 270:5950-5953(1995).			
CC	-1- FUNCTION: ORPHAN RECEPTOR.			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.			
CC	-1- PTM: ACTIVATED BY PROTEOLYTIC CLEAVAGE OF ITS EXTRACELLULAR N-TERMINUS.			
CC	-1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; Z48043; G663021; -			
DR	GCRDB; GCR_1727; -			
DR	MGI; MGI:101910; GPCR11.			
DR	PROSITE; PS00237; G-PROTEIN_RECEPTOR; FALSE_NEG.			
DR	PFAM; PF00001; 7tm_1; 1.			
DR	HSSP; P34996; 1DDD.			
DR	G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.			
FT	SIGNAL	1 25		
FT	PROPEP	26 38		
FT	-----			
FT	CHAIN	39 399		
FT	DOMAIN	39 77		
FT	TRANSNEM	78 103		
FT	DOMAIN	104 112		
FT	TRANSNEM	113 132		
FT	DOMAIN	133 151		
FT	TRANSNEM	152 173		
FT	DOMAIN	174 192		
FT	TRANSNEM	193 213		
FT	DOMAIN	214 243		
FT	TRANSNEM	244 262		
FT	DOMAIN	263 287		
FT	TRANSNEM	288 311		
FT	DOMAIN	311 325		
FT	TRANSNEM	326 349		
FT	DOMAIN	350 399		
FT	DISULFID	150 228		
FT	CARBOHYD	33 33		
FT	CARBOHYD	224 224		
FT	SEQUENCE	399 AA; 44752 MW; D4EA5805 CRC32;		

Query Match

26.6%; Score 530; DB 1; Length 399;

Best Local Similarity 34.1%; Pred. No. 1e-27;
Matches 131; Conservative 68; Mismatches 141; Indels 44; Gaps 11;

QY 13 GFSLSGGT-PSVYDESGTGGDDSTPSILPAPRGYQGVQVANDSDTLELPDSSRALL 71
Db 37 GRSIGRLETQPPi-----TGKG-----VPVEPGF-----SIDFESASIL 71

QY 72 LGWVTRLPALYGLVVLVGLPANGIALWV-LATQAPRLPSTMLMNLATADLLALALP 130
Db 72 TGLTTPVLPVVIIVFVIGLPSNGMALWFLTRTKKHPAVIYMANLADLLSVIWP 131

QY 131 PRIAYHLRQRPFGAACLATAALYGHMYGVSLLAAVSLDRYALVHPL---RARAL 187
Db 132 LKISYHLGNWVYIGALCKVLGIFGYGMYCSILFMTCLSVQYVWVIVNPMGHPKKA- 191

QY 188 RGRRLALGCMMAWMAALALPLTQROTFLARSDRVLCHDALPLDAQASHWQPAFTC 247
Db 191 ---NIAGVSLAIWLLIFLVTIPLYVMKOTIYPALNITTHDVLPEEVLVGMDFNYLS 247

QY 248 LALLGCF-PLLAMLCYGATLHTLAAG-----RRYGHALRLTAVVLASAVAFVPSN 300
Db 248 LA-IGVFLPALLTASAYVLMIKTLASSAMDEHSEKKRRRAIRLIITVLMYFICFAPS 306

QY 301 LLLLHYDPSAMGNLYGAVPSLALSTLNSCVDPFYIYVSAEFRKVRAGLFQRS- 360
Db 307 LLLVHYFLIKTQSHVYALVLCVLTSLNSCIDPFYVYFVSKDFRDHARNALICRSV 366

QY 360 -PGDTVASKASAGSGSGMGTHSS 382
Db 367 RVNRMQISLSSNFKRSGSYSS 390

RESULT 2
ID PAR2_RAT STANDARD; PRT; 397 AA.
AC Q63645;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PROTEINASE ACTIVATED RECEPTOR 2 PRECURSOR (PAR-2).
GN PAR2.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP STRAIN-SPRAGUE-DAWLEY; TISSUE=INTESTINE, AND KIDNEY;
RX MEDLINE; 96358009.
RA SAFFEDINE M., AL-ANI B., CHENG C.H., WANG L., HOLLENBERG M.D.;
RT "Rat proteinase-activated receptor-2 (PAR-2): cDNA sequence and
RT activity of receptor-derived peptides in gastric and vascular
RT tissue.";
RL BR. J. PHARMACOL. 118:521-530(1996).
CC - FUNCTION: ORPHAN RECEPTOR.
CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC - PTM: ACTIVATED BY PROTEOLYTIC CLEAVAGE OF ITS EXTRACELLULAR N-
CC TERMINUS.
CC - SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U61373; G1480688; -
CC PROSITE; PS00237; G-PROTEIN_RECEPTOR; FALSE_NEG.
CC PFAM; PF00001; 7tm_1; 1.
CC HSP; P34996; 1DDO.
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 25 POTENTIAL.

FT PROPEP 26 36 REMOVED FOR RECEPTOR ACTIVATION
FT CHAIN 37 397 (BY SIMILARITY).
FT DOMAIN 37 75 PROTEINASE ACTIVATED RECEPTOR 2.
FT TRANSEM 76 101 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 102 110 1 (POTENTIAL).
FT TRANSEM 111 130 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 131 149 2 (POTENTIAL).
FT TRANSEM 150 171 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 172 190 3 (POTENTIAL).
FT TRANSEM 191 211 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 212 241 4 (POTENTIAL).
FT TRANSEM 242 260 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 261 285 5 (POTENTIAL).
FT TRANSEM 286 308 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 309 323 6 (POTENTIAL).
FT TRANSEM 324 347 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 348 397 7 (POTENTIAL).
FT DISULFID 148 226 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 23 23 BY SIMILARITY.
FT CARBOHYD 222 222 POTENTIAL.
SQ SEQUENCE 397 AA; 44440 MW; D59C4FEC CRC32;

Query Match 26.3%; Score 524; DB 1; Length 397;
Best Local Similarity 34.6%; Pred. No. 2.4e-27;
Matches 124; Conservative 62; Mismatches 132; Indels 40; Gaps 9;

QY 13 GFSLSGGTQPSVYDESGTGGDDSTPSILPAPRGYQGVQVANDSDTLELPDSSRALLL 72
Db 35 GRSIGRLDTTPPi-----TGKG-----APVEPGF-----SVDEFASVLT 70

QY 73 GWVTRLPALYGLVVLVGLPANGIALWV-LATQAPRL-PSTMLMNLATADLLALALPP 131
Db 71 GKLITVFLPVYIIVFVIGLPSNGMALWVFFTRKKKHPAVIYMANLADLLSVIWEPL 130

QY 132 RIAYHLRQRPFGAACLATAALYGHMYGVSLLAAVSLDRYALVHPL---RARALR 188
Db 131 KISYHLGNWVYIGALCKVLGIFGYGMYCSILFMTCLSVQYVWVIVNPMGHSKRKA-- 189

QY 189 GRLALGCMMAWMAALALPLTQROTFLARSDRVLCHDALPLDAQASHWQPAFTCL 248
Db 189 ---NIAGVSLAIWLLIFLVTIPLYVMKOTIYPALNITTHDVLPEEVLVGMDFSYFUSL 246

QY 249 ALLGCF-PLLAMLCYGATLHTLAAG-----RRYGHALRLTAVVLASAVAFVPSNL 301
Db 247 A-IGVFLPALLTASAYVLMIKTLASSAMDEHSEKKRRRAIRLIITVLSMYFICFAPS 305

QY 302 LLLLHYDPSAMGNLYGAVPSLALSTLNSCVDPFYIYVSAEFRKVRAGLFQRS 359
Db 306 LLLVHYFLIKTQSHVYALVLCVLTSLNSCIDPFYVYFVSKDFRDQARNALLCRS 363

RESULT 3
ID PAR3_HUMAN STANDARD; PRT; 374 AA.
AC Q00254;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROTEINASE ACTIVATED RECEPTOR 3 PRECURSOR (PAR-3) (THROMBIN RECEPTOR
DE 2).
GN PAR3.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97242411.
RA ISHIMURA H., CONNOLLY A.J., ZENG D., KAHN M.L., ZHENG Y.W.,
RA TIMMONS C., TRAM T., COUGHLIN S.R.;
RT "Protease-activated receptor 3 is a second thrombin receptor in
RT humans.";


```
RESULT 6
THRR_XENLA STANDARD; PRT; 420 AA.
AC P47749;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE THROMBIN RECEPTOR PRECURSOR.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPODEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94195429.
RA GERSTEN R.E., CHEN J., ISHII M., ISHII K., NANEVICZ T.,
RA TURCK C.W., VU T.-K.H., COUGHLIN S.R.;
RT "Specificity of the thrombin receptor for agonist peptide is defined
RT by its extracellular surface.";
RL NATURE 368:648-651(1994).
CC -1- FUNCTION: RECEPTOR FOR ACTIVATED THROMBIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- PTM: IT IS THOUGHT THAT CLEAVAGE AFTER AA 42 BY THROMBIN LEADS TO
CC ACTIVATION OF THE RECEPTOR. THE NEW AMINO TERMINUS FUNCTIONS AS
CC A TETHERED LIGAND AND ACTIVATES THE RECEPTOR.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U09632; G495198;
CC GCRDB: GCR_1038;
CC PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.
CC PRAM; PF00001; 7tm_1; 1.
CC G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
CC BLOOD COAGULATION.
CC SIGNAL 1 20
CC PROPEP 21 42
CC CHAIN 43 420
CC DOMAIN 43 101
CC TRANSMEM 102 127
CC DOMAIN 128 136
CC TRANSMEM 137 156
CC DOMAIN 157 175
CC TRANSMEM 176 197
CC DOMAIN 198 217
CC TRANSMEM 218 238
CC DOMAIN 239 267
CC TRANSMEM 268 287
CC DOMAIN 288 310
CC TRANSMEM 311 330
CC DOMAIN 334 345
CC TRANSMEM 346 369
CC DOMAIN 370 420
CC SITE 42 43
CC DISULFID 174 253
CC CARBOHYD 38 38
CC CARBOHYD 86 86
CC MUTAGEN 42 42
CC K-2A: LACK OF ACTIVATION BY THROMBIN.
CC SEQUENCE 420 AA; 47435 MW; D21E8698 CRC32;

Query Match 24.5%; Score 488; DB 1; Length 420;
Best Local Similarity 34.9%; Pred. No. 5.4e-25;
Matches 124; Conservative 60; Mismatches 143; Indels 28; Gaps 8;

QY 27 DESGSGGDDSTFSLPAPRGVQGVCAVNDSDLELPDSSRALLGWVTRLVPLYGL 86
```

```
Db 63 DESGE-GSGDQA-----PVSRSARKFIRNITKEAOYLSSQ-----WL-TKVPVSLYTV 110
QY 87 VLVVGLPANGALWV-LATQAPRLPSTMLMNLATALLALALPPRIAYHLRGQRWPF 145
Db 111 VFIVGLPLNLLAIIIFLEKMKVKPAVVMYMLNLAIDVFFSVLPFKIAYHLSGNDWLF 170
QY 146 EACRATATAALYGHMYGVSLLLAASLDRLALVHLPARALGRRLALGLCHNAWMAA 205
Db 171 PGMCRIVTAIFYCNMYCSVLLIASISVDRFLAVVYPMHLSLWRTMSRAYNACSFIMLSI 230
QY 206 ALALPLTQROTFLARSDRVLCCHDALPDAQASHQWPAFTCLALLGCFPLLLAMLICYG 265
Db 231 ASIPLLVTEQTKIPRLDTTCHVDVLDKDLDFIYFFSSFCLLFFFFFIITICYI 290
QY 266 ATLHTLAAS-----GRRYGHALRLTAVVLASAVAFFVPSNLLLLHYSDPSAMGNLYGA 321
Db 291 GIIRSLSSSIENSCKTKRALFLAVVVLVCFIICFGPTNVLFTHVLQEADEF---LYFA 347
QY 322 YVPSLALSTLNSVDPIFYIYVSAEFRDKVRAGLFORSPOGDTVASKASAEAGSGRG 376
Db 348 YILSACVSGVSCCLDPLIYYIYSSQCRVLYSLCCR-----KVSEPGSGSTG 394

RESULT 7
THRR_PAPHA STANDARD; PRT; 425 AA.
AC P56488;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE THROMBIN RECEPTOR PRECURSOR.
GN F2R OR PAR1 OR BTHR12.
OS PAPIO HAMADRYAS (HAMADRYAS BABOON).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; PAPIO.
RN [1]
RP SEQUENCE FROM N.A.
RA SHOJI M., HAYZER D.J., HANSON S.R.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: RECEPTOR FOR ACTIVATED THROMBIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- PTM: IT IS THOUGHT THAT CLEAVAGE AFTER AA 41 BY THROMBIN LEADS TO
CC ACTIVATION OF THE RECEPTOR. THE NEW AMINO TERMINUS FUNCTIONS AS
CC A TETHERED LIGAND AND ACTIVATES THE RECEPTOR.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF028727; G3605877;
CC GCRDB: GCR_2523;
CC PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.
CC G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
CC BLOOD COAGULATION.
CC SIGNAL 1 26
CC PROPEP 27 41
CC CHAIN 42 425
CC DOMAIN 42 102
CC TRANSMEM 103 128
CC DOMAIN 129 137
CC TRANSMEM 138 157
CC DOMAIN 158 176
CC TRANSMEM 177 198
CC DOMAIN 199 218
CC TRANSMEM 219 239
CC DOMAIN 240 268
CC TRANSMEM 269 288
CC REMOVED FOR RECEPTOR ACTIVATION.
CC THROMBIN RECEPTOR.
CC EXTRACELLULAR (POTENTIAL).
CC 1 (POTENTIAL).
CC CYTOPLASMIC (POTENTIAL).
CC 2 (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC 3 (POTENTIAL).
CC CYTOPLASMIC (POTENTIAL).
CC 4 (POTENTIAL).
CC EXTRACELLULAR (POTENTIAL).
CC 5 (POTENTIAL).
```

```
FT DOMAIN 289 311 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 312 334 6 (POTENTIAL).
FT DOMAIN 335 350 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 351 374 7 (POTENTIAL).
FT DOMAIN 375 425 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 35 35 POTENTIAL.
FT CARBOHYD 62 62 POTENTIAL.
FT CARBOHYD 75 75 POTENTIAL.
FT CARBOHYD 250 250 POTENTIAL.
FT CARBOHYD 259 259 POTENTIAL.
FT SITE 41 42 CLEAVAGE (BY THROMBIN).
FT DOMAIN 57 60 ASP/GLU-RICH (ACIDIC).
FT DISULFID 175 254 BY SIMILARITY.
SQ SEQUENCE 425 AA; 47253 MW; 9E709862 CRC32;
```

Query Match 24.1%; Score 480; DB 1; Length 425;
Best Local Similarity 33.9%; Pred. No. 1.8e-24;
Matches 117; Conservative 70; Mismatches 144; Indels 14; Gaps 5;

```
QY 53 VCANDSDTLEP-----DSSRALLGWPTRLPALYGLVLPANGALAV-LAT 104
Db 72 VSINKSPLOKPLPAFISDAGYLTSSWL-TLFVPSVITGVVSLPVMIAVVFILK 130
QY 105 QAPRLPSTMLMNLATADLLALPPRIAYHLRGQRPFGEAACRLATAALYGHMYSV 164
Db 131 MKVKPAPVYMLHLATADVLSVLPFKISYLSGSDMQFSGELCFVTAAYFCNMYASI 190
QY 165 LLLAAVSLDRYALVHLPRARALGRRLALGICMAWMAALALPLTLQRTFLARS 224
Db 191 LLMTVISIDRLAVVPMQSLWRTLGRASFCTCLAIWALAIAGVVPVLLKEQTIQV 250
QY 225 RVLCHDALPDAQASHQWPAFTCLALLGCFPLLAMLYCGATLHTLA---ASGRYGH 280
Db 251 ITCHDVNLNETLEGYYAYFSAFSAFVFFVLIISTVCYVSIIRCLSSSTVANSKSR 310
QY 281 ALRLAVLASAVAFVPSNLLLLHYDPS-PSAWGNLYGAYVPSLALSTLNSCVDPFI 339
Db 311 ALFLSAAVFCIFICGPTNILLIAHYSPLSTSTTEAAYFAYLLVCVSSISCCIDPLI 370
QY 340 YYYVSAERDKVRAGLFQSPGDTVASKASAGSGRGMGTHSSLL 384
Db 371 YYYASQRYVYIILCKESSDPSNSSGQIMASKMDTCSNLL 415
```

```
RESULT 8
ID THRR_HUMAN STANDARD; PRT: 425 AA.
AC P25116;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE THROMBIN RECEPTOR PRECURSOR.
GN F2R OR PAR1 OR TR.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91168254.
RA VU T.-K.H., HUNG D.T., WHEATON V.I., COUGHLIN S.R.;
RT "Molecular cloning of a functional thrombin receptor reveals a novel
RT proteolytic mechanism of receptor activation.";
RL CELL 64:1057-1068(1991).
CC -!- FUNCTION: RECEPTOR FOR ACTIVATED THROMBIN.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: PLATELETS AND VASCULAR ENDOTHELIAL CELLS.
CC -!- PM: IT IS THOUGHT THAT CLEAVAGE AFTER AA 41 BY THROMBIN LEADS TO
CC ACTIVATION OF THE RECEPTOR. THE NEW AMINO TERMINUS FUNCTIONS AS
CC A TETHERED LIGAND AND ACTIVATES THE RECEPTOR.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M62424; G339677; -.
DR PIR; A37912; A37912.
DR GCRDB; GCR_0088; -.
DR MIN; 187930; -.
DR PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.
DR PFAM; PF00001; 7tm1.1.
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
KW BLOOD COAGULATION.
FT SIGNAL 1 26 POTENTIAL.
FT PROPEP 27 41 REMOVED FOR RECEPTOR ACTIVATION.
FT CHAIN 42 425 THROMBIN RECEPTOR.
FT DOMAIN 42 102 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 103 128 1 (POTENTIAL).
FT DOMAIN 129 137 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 138 157 2 (POTENTIAL).
FT DOMAIN 158 176 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 177 198 3 (POTENTIAL).
FT DOMAIN 199 218 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 219 239 4 (POTENTIAL).
FT DOMAIN 240 268 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 269 288 5 (POTENTIAL).
FT DOMAIN 289 311 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 312 334 6 (POTENTIAL).
FT DOMAIN 335 350 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 351 374 7 (POTENTIAL).
FT DOMAIN 375 425 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 35 35 POTENTIAL.
FT CARBOHYD 62 62 POTENTIAL.
FT CARBOHYD 75 75 POTENTIAL.
FT CARBOHYD 250 250 POTENTIAL.
FT CARBOHYD 259 259 POTENTIAL.
FT SITE 41 42 CLEAVAGE (BY THROMBIN).
FT DOMAIN 57 60 ASP/GLU-RICH (ACIDIC).
FT DISULFID 175 254 BY SIMILARITY.
SQ SEQUENCE 425 AA; 47410 MW; E9A485AE CRC32;
```

```
Query Match 23.8%; Score 473; DB 1; Length 425;
Best Local Similarity 32.1%; Pred. No. 5e-24;
Matches 133; Conservative 76; Mismatches 169; Indels 36; Gaps 10;

QY 4 RLLMLPLVGLFSLSG---GTQTPSVYDESGTGGDDSTFSLPAP-----RG 48
Db 5 RLLL--VAACPSLCGPLLARSARRPESKATNATLDRSFLLRPNNDKVEPFWEDEKN 62
QY 49 YPG-----QVCANDSDYL--ELP-----DSSRALLGWPTRLPALYGLVLPANG 96
Db 63 ESGTEYRLVSIKNSPLQKLPFAFISDAGYLTSSWL-TLFVPSVITGVVSLPLNI 121
QY 97 LALWV-LATQAPRLPSTMLMNLATADLLALPPRIAYHLRGQRPFGEAACRLATAA 155
Db 122 MAIVVFILKMKVKPAPVYMLHLATADVLSVLPFKISYLSGSDMQFSGELCFVTA 181
QY 156 LYGHMYSVLLAAVSLDRYALVHLPRARALGRRLALGICMAWMAALALPLTLQ 215
Db 182 FYCNMYASILLMTVISIDRLAVVPMQSLWRTLGRASFCTCLAIWALAIAGVVP 241
QY 216 QTFRLASDRVLCDALPDAQASHQWPAFTCLALLGCFPLLAMLYCGATLHTLA 275
Db 242 QTIQVPLNITTDHVLNETLEGYYAYFSAFSAFVFFVLIISTVCYVSIIRCLSS 301
QY 276 ----RRYGHALRLAVLASAVAFVPSNLLLLHYDPS-PSAWGNLYGAYVPSLA 330
Db 302 VANSKSRALFLSAVFCIFICGPTNILLIAHYSPLSTSTTEAAYFAYLLVCVSS 361
QY 331 LNSCVDPFIYYVSAERDKVRAGLFQSPGDTVASKASAGSGRGMGTHSSLL 384
```


DE THROMBIN RECEPTOR PRECURSOR.
GN F2R OR PAR1.
OS CRICETULUS LONGICAUDATUS (LONG-TAILED HAMSTER) (CHINESE HAMSTER).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; CRICETINAE; CRICETULUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE: 91348247.
RA RASMUSSEN U B., VOURET-CRAVIARI V., JALLAT S., SCHLESINGER Y.,
RA PAGES G., PAVIRANI A., LECOQ J.P., FOURSEGUER J.,
RA OBERGHEEN-SCHILLING E.;
RT "cDNA cloning and expression of a hamster alpha-thrombin receptor
RT coupled to Ca2+ mobilization.";
RL FEBS LETT. 288:123-128(1991).
RN [2]
RP SEQUENCE OF 42-428 FROM N.A.
RX MEDLINE: 96028007.
RA HARTMANN T., GRACE M.B., BUZARD G.S., RUOSS S.J.;
RT "Thrombin receptor polymorphism in Chinese hamster.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 215:974-980(1995).
CC -1- FUNCTION: RECEPTOR FOR ACTIVATED THROMBIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- PTM: IT IS THOUGHT THAT CLEAVAGE AFTER AA 41 BY THROMBIN LEADS TO
CC ACTIVATION OF THE RECEPTOR. THE NEW AMINO TERMINUS FUNCTIONS AS
CC A TETHERED LIGAND AND ACTIVATES THE RECEPTOR.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X61958; G49538; -.
CC DR EMBL: U34047; G1000286; -.
CC DR PIR: S17148; S17148.
CC DR GCRDB: GCR_0166; -.
CC DR GCRDB: GCR_0432; -.
CC DR GCRDB: GCR_1173; -.
CC DR PROSITE: PS00237; G_PROTEIN_RECEPTOR; 1.
CC DR PFAM: PF00001; 7tm_1; 1.
CC DR HSSP: P34996; 1DDD.
CC KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
CC BLOOD COAGULATION; POLYMORPHISM.
FT SIGNAL 1 ?
FT PROPEP 1 41
FT REMOVED FOR RECEPTOR ACTIVATION
FT (BY SIMILARITY).
FT THROMBIN RECEPTOR.
FT CHAIN 42 428
FT DOMAIN 42 105
FT EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 106 131
FT DOMAIN 132 140
FT CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 141 160
FT DOMAIN 161 179
FT EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 180 201
FT DOMAIN 202 221
FT CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 222 242
FT DOMAIN 243 271
FT EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 272 291
FT DOMAIN 292 314
FT CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 315 337
FT DOMAIN 338 352
FT EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 353 377
FT DOMAIN 378 428
FT CYTOPLASMIC (POTENTIAL).
FT SLITE 41 42
FT CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT DOMAIN 60 63
FT DISULFID 178 257
FT CARBOHYD 65 65
FT CARBOHYD 78 78
FT CARBOHYD 253 253
FT CARBOHYD 262 262

FT VARIANT 84 84 MISSING (IN AN ALLELE).
FT CONFLICT 384 384 H -> T (IN REF. 1).
SQ SEQUENCE 428 AA; 47602 MW; BC709D4C CRC32;

Query Match 23.0%; Score 458.5; DB 1; Length 428;
Best Local Similarity 32.4%; Pred. No. 4.4e-23;
Matches 126; Conservative 69; Mismatches 135; Indels 59; Gaps 14;

QY 4 RLLPLVILGFSLSG---GTQTPSYVDESGST-----DSSRALLLGWVTRL--VPALYGLVL 36
DB 5 RLLL--VAAGSLGCLPLSSRPVQPESEMTDATVNPFRFLRNPGEFTFELPLGDEE 62
QY 37 DSTPSILPAPGYGQVCANDSDTELP-----IYLNKSHPPAPLAFISEDAGSYLTSPWL--RLIPSYVTFV 88
DB 63 EKNESTLPEGRA-----IYLNKSHPPAPLAFISEDAGSYLTSPWL--RLIPSYVTFV 116
QY 89 VVGLPANGALWVLATQAP--RLPSTMLLMNLTADLLALALPPRIAYHLRGQRPFGEA 147
DB 117 VVSLPLNLATAVFLKMKVKPAVVMHLAMADVLSVLPKISYFSGSDWQFGSG 176
QY 148 ACRLATAALGHMYGVSLLLLAAVSLDRYLALVHLRRLRRLALGLCMAAAL 207
DB 177 MCRFAATAFYCNMYASIMLMTVISIDRELVAVYPIQSLSWRTGLRANFTCLVIVWMAIMG 236
QY 208 ALPLTQRTFLRARSRLVCHDALP---LDAQASHMQPAFTCLALLGCFPLLLAMLICY 264
DB 237 VVPLLLKEQTTRVGLNITTCHDVNLQGFYSYFSAFSAVFFL---VPLIISTICY 293
QY 265 GATLHTLAASG-----RRYGHALRTAVVLASAVAFFVPSNLLLLLHY---SDPSPSANGN 317
DB 294 MSIIICLSSSVANRKSRAFLFLSAAVFCVFCVFGPTNVLIMHYLLSD-SP-ATEK 351
QY 318 LYGAVVPSIALSTLNSCVDPETIYYVSAE 346
DB 352 AYFALLCVSVSSVSCCIDPLIYYIASSE 380

RESULT 11
THRR_RAT
ID THRR_RAT STANDARD; PRT; 432 AA.
AC P26824;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE THROMBIN RECEPTOR PRECURSOR.
GN F2R OR PAR1.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ARTICULAR SMOOTH MUSCLE;
RX MEDLINE: 92381002.
RA ZHONG C., HAYZER D.J., CORSEN M.A., WICK K., RUNGE M.S.;
RT "Molecular cloning of the rat vascular smooth muscle thrombin
RT receptor. Evidence for in vitro regulation by basic fibroblast growth
RT factor.";
RL J. BIOL. CHEM. 267:16975-16979(1992).
CC -1- FUNCTION: RECEPTOR FOR ACTIVATED THROMBIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- PTM: IT IS THOUGHT THAT CLEAVAGE AFTER AA 45 BY THROMBIN LEADS TO
CC ACTIVATION OF THE RECEPTOR. THE NEW AMINO TERMINUS FUNCTIONS AS
CC A TETHERED LIGAND AND ACTIVATES THE RECEPTOR.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

```
CC EMBL; M81642; G207466; -.
DR PIR; A43448; A43448.
DR GCRDB; GCR_0263; -.
DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR PFAM; PF00001; 7tm_1; 1.
DR HSP; P34996; 1DDD.
DR G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
KW BLOOD COAGULATION.
FT SIGNAL 1 ?
FT PROPEP ? 45
FT CHAIN 46 432
FT DOMAIN 46 109
FT TRANSNEM 110 135
FT DOMAIN 136 144
FT TRANSNEM 145 164
FT DOMAIN 165 183
FT TRANSNEM 184 205
FT DOMAIN 206 225
FT TRANSNEM 226 246
FT DOMAIN 247 275
FT TRANSNEM 276 295
FT DOMAIN 296 318
FT TRANSNEM 319 341
FT DOMAIN 342 357
FT TRANSNEM 358 381
FT DOMAIN 382 432
FT SITE 45 46
FT DOMAIN 64 68
FT DISULFID 182 261
FT CARBOHYD 69 82
FT CARBOHYD 82 257
FT CARBOHYD 257 266
FT CARBOHYD 266 266
FT SEQUENCE 432 AA; 48280 MW; 758B2DA8 CRC32;

Query Match 22.9%; Score 456.5; DB 1; Length 432;
Best Local Similarity 30.3%; Pred. No. 5.9e-23;
Matches 132; Conservative 79; Mismatches 155; Indels 67; Gaps 15;

QY 4 RLLWPLVLGSLG---GTQTPSVYDESGTGGDDSTPSILPAPRGYQGVCAANDS-D 59
Db 5 RLLWPLVLGSLG---GTQTPSVYDESGTGGDDSTPSILPAPRGYQGVCAANDS-D 59
QY 60 TLE-----LPDSSRALLG-----WVPTLVPA 82
Db 56 TFEQFPLGDEEKESIPLEGRAVYLNKSRPPMPFPFISEDASGYLTSPWL-TLIPS 114
QY 83 LYGLVVLVGLPANGALWVLAQAP-RLPSTMLMNLATADLLALALPPIAYHLRQOR 141
Db 115 VYTFVFLVSLPLNLAIAVFFVRMKKPAVVMYMLAMADVFLVSVLPFKISYFSGTD 174
QY 142 WPFGEACRLATAALYGHMYGVLLLAASLDRLVALVHLPLARALGRRLALGLCAAW 201
Db 175 WQFSGMCREATAACYCMYASIMLMTVTSIDRELAIVYQSLSWRTLGRANFTCVIWI 234
QY 202 LMAAALALPTLQRTFLRARSRLVCHDALP---LDAQASHMQPFTCLALLGCLPL 258
Db 235 VMAINGVVPVLLKQTTQVPLNITTCHVNLNELLHGFSYYSFAIFFL---VPLI 291
QY 259 AMLLCYGATLHTLAASG-----RRYGHALRLTAVVLASAVAFVPSNLLLLHY---SDPS 311
Db 292 ISTVCYTSIIRCLSSAVANRKSRLAFLSAVFCIFVCGFTNVLIVHLLLS-D-S 350
QY 312 PSANGNLYGAVPSLALSTLNCVDYFYIYVYSAEERDKVRAGLFQSPGDTVASKASAE 371
Db 351 PGT-ETAFYALGCVVTSVASCIDPLIYYIYASSECOKHLYSILCCRSDNSCNSGTG 409
QY 372 GGSRGMTGTHSLL 384
Db 410 LMPKMDTCSSHL 422
```

RESULT 12
P2Y5_CHICK
ID P2Y5_CHICK STANDARD; PRT; 308 AA.
AC P32250;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE P2Y PURINOCEPTOR 5 (P2Y5) (PURINERGIC RECEPTOR 5) (6H1).
GN P2Y5.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
[1]
RN RP SEQUENCE FROM N.A.
RC MEDLINE; 93329058.
RX KAPLAN M.H., SMITH D.I., SUNDICK R.S.;
RA "Identification of a G protein coupled receptor induced in activated
RT T cells.";
RL J. IMMUNOL. 151:628-636(1993).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: INDUCED IN ACTIVATED T-CELLS.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; L06109; G304384; -.
DR GCRDB; GCR_0689; -.
DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR PFAM; PF00001; 7tm_1; 1.
DR HSP; P34996; 1DDD.
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; LIPOPROTEIN;
KW PALMITATE.
FT DOMAIN 1 16 EXTRACELLULAR (POTENTIAL).
FT TRANSNEM 17 43 1 (POTENTIAL).
FT DOMAIN 44 52 CYTOPLASMIC (POTENTIAL).
FT TRANSNEM 53 76 2 (POTENTIAL).
FT DOMAIN 77 89 EXTRACELLULAR (POTENTIAL).
FT TRANSNEM 90 109 3 (POTENTIAL).
FT DOMAIN 110 130 4 (POTENTIAL).
FT TRANSNEM 131 151 4 (POTENTIAL).
FT DOMAIN 152 178 EXTRACELLULAR (POTENTIAL).
FT TRANSNEM 179 206 5 (POTENTIAL).
FT DOMAIN 207 224 CYTOPLASMIC (POTENTIAL).
FT TRANSNEM 225 250 6 (POTENTIAL).
FT DOMAIN 251 269 EXTRACELLULAR (POTENTIAL).
FT TRANSNEM 270 289 7 (POTENTIAL).
FT DOMAIN 290 308 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 5 5 POTENTIAL.
FT LIPID 281 281 PALMITATE (BY SIMILARITY).
FT DISULFID 86 165 POTENTIAL.
FT SEQUENCE 308 AA; 35597 MW; B2DFEAD4 CRC32;

Query Match 18.7%; Score 372.5; DB 1; Length 308;
Best Local Similarity 30.3%; Pred. No. 1.1e-17;
Matches 91; Conservative 57; Mismatches 129; Indels 23; Gaps 7;

QY 58 SDTLELPSDRALLGHWPTPLVPALYGLVVLVGLPANGALWVLA-TQAPRLPSTMLLM 116
Db 3 SSNCSTEDSFYTLGCV-----FSMVFVLGLIANCAVAYIFTFTLVKRNFTTYML 54
QY 117 NLATADLLALALPPIAYHLRQORWPFGEACRLATAALYGHMYGVLLLAASLDRLVAL 176
Db 55 NLASDLLLFTVLPFRI-YFVYVNRWPFGEACRLATAALYGHMYGVLLLAASLDRLVAL 113

Query Match 18.6%; Score 371; DB 1; Length 328;

7:

[illegible]

- |- FUNCTION: RECEPTOR FOR EXTRACELLULAR UDP > UTP > ATP. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
- |- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
- |- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

ENBL; X97058; E236011; -.
ENBL; U52464; G1407633; -.
ENBL; AF007892; G2258422; -.
ENBL; AF007891; G2258420; -.
MIN; 602451; -.
PROSITE; PS00237; G_PROTEIN_RECEPTOR; FALSE_NEG.
PFAM; PF00001; 7tm_1; 1.
HSSP; P34996; 1DDD.

```

CHANG K., HANAKURA K., KUMADA M., TAKUWA I.;
RT "Molecular cloning and functional analysis of a novel P2 nucleotide
RT receptor.";
RL J. BIOL. CHEM. 270:26152-26158(1995).
CC -|- FUNCTION: RECEPTOR FOR EXTRACELLULAR UTP > ADP = 2-METHYLTHIO-A
CC > ADP-BETA-S > ATP = ATP-GAMMA-S. THE ACTIVITY OF THIS RECEPTOR


```

Db 125 ADRFLLVFNPIWCONFGRGAGLAWIACAVAW-----GLALLLTIPSLYRVVREYFPPKVL 180
QY 228 C-----HDAFLPLDAQASHWQPAFTCLALLGCFPLLAMLLCYG-ATLHTLAASGRRYGHA 281
Db 181 CGVDHGHDKRRERAVA-----IARLVLGFWYPLTLTMCYTFLLLRWSSRRATRSTKT 233
QY 282 LRLTAVVLASAVAFVPSNLL-LLHYSPPSPSANGNLYGAVVPSIALS--TLNSCVDPF 338
Db 234 LKVVAVVASFFIFWLPYQVTGMMMSFLEPSSPTF--LLKKLDSLCISFAYINCCINPI 291
QY 339 IYYVVSABERDKVRAGL-----FORSPGDTVASK 367
Db 292 IYVAGOGFQGRLRKSLPSSLRLNVLTEESMVRESKSFTRSTVDVTMAQK 339

```

Search completed: October 29, 1999, 15:52:27
 Job time: 17104 sec

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	530	26.6	399	1	US-08-476-000-61	Sequence 61, Appl	
2	530	26.6	399	2	US-08-472-840-61	Sequence 61, Appl	
3	530	26.6	399	2	US-08-476-976-61	Sequence 61, Appl	
4	528	26.5	395	1	US-08-097-938-2	Sequence 2, Appl	
5	528	26.5	395	1	US-08-476-000-2	Sequence 2, Appl	
6	528	26.5	395	2	US-08-472-840-2	Sequence 2, Appl	
7	528	26.5	395	2	US-08-476-976-2	Sequence 2, Appl	
8	525	26.4	395	1	US-08-097-938-5	Sequence 5, Appl	
9	525	26.4	395	1	US-08-476-000-5	Sequence 5, Appl	
10	525	26.4	395	2	US-08-472-840-5	Sequence 5, Appl	
11	525	26.4	395	2	US-08-476-976-5	Sequence 5, Appl	
12	521.5	26.2	408	2	US-08-742-440A-6	Sequence 6, Appl	
13	510	25.6	398	1	US-08-097-938-6	Sequence 6, Appl	
14	510	25.6	398	1	US-08-476-000-6	Sequence 6, Appl	
15	510	25.6	398	2	US-08-472-840-6	Sequence 6, Appl	
16	510	25.6	398	2	US-08-476-976-6	Sequence 6, Appl	
17	508	25.5	398	1	US-08-097-938-4	Sequence 4, Appl	
18	508	25.5	398	1	US-08-476-000-4	Sequence 4, Appl	
19	508	25.5	398	2	US-08-472-840-4	Sequence 4, Appl	
20	508	25.5	398	2	US-08-476-976-4	Sequence 4, Appl	
21	507.5	25.5	407	2	US-08-742-440A-3	Sequence 3, Appl	
22	505.5	25.4	394	2	US-08-742-440A-8	Sequence 8, Appl	
23	494	24.8	397	1	US-08-476-000-63	Sequence 63, Appl	
24	494	24.8	397	2	US-08-472-840-63	Sequence 63, Appl	
25	494	24.8	397	2	US-08-476-976-63	Sequence 63, Appl	
26	473	23.8	425	1	US-07-657-7698-69	Sequence 69, Appl	
27	473	23.8	425	1	US-08-097-938-7	Sequence 7, Appl	
28	473	23.8	425	1	US-07-789-184-220	Sequence 220, App	
29	473	23.8	425	1	US-08-476-000-7	Sequence 7, Appl	
30	473	23.8	425	2	US-08-475-263-220	Sequence 220, App	
31	473	23.8	425	2	US-08-472-840-7	Sequence 7, Appl	
32	473	23.8	425	2	US-08-485-886-220	Sequence 220, App	
33	473	23.8	425	2	US-08-477-362-220	Sequence 220, App	
34	473	23.8	425	2	US-08-477-134-220	Sequence 220, App	
35	473	23.8	425	2	US-08-911-320A-3	Sequence 3, Appl	
36	473	23.8	425	2	US-08-476-976-7	Sequence 7, Appl	
37	473	23.8	425	2	US-08-742-440A-7	Sequence 7, Appl	
38	472	23.7	425	1	US-08-313-553-13	Sequence 13, Appl	
39	448.5	22.5	357	3	PCR-US95-07180-3	Sequence 3, Appl	

```

Db 72 TGKLTTFVPLVYIIVFVIGLPSNGMALWIFLFTRTKKKKHPAVIYMANLALADLLSVIWF 131
Qy 131 PRIAYHRLGQRWPEGEACRLATRAALYGHMVGSVLLAAVSLDRYALVHPL---RARAL 187
Db 132 LKISVHLGNWWTYGEALCKYKLIQFFTGNTWCSLFTWCSVQRYWTVIAPMGHPRKA- 191
Qy 188 RGRRLALGLCMAAALMALPLTLQRTFRLARSDRVLC HDALPLDAQASHWQPAFTC 247
Db 191 --NIAVGSALIMLLIFLTVPLVVMKQTIYPALNITTC HDVLPPEVLVGD MFN FLS 247
Qy 248 LALLGCFL-PLIALLMCCYGATLHTLAAG-----RRYGHALRLTAVVLASAVFAFFVPSN 300
Db 248 LA-IGVFLPALLTASAYVLMIKTLRSAMDHSEKSRQRAIRLIITVLAMYFTCFEAPS 306
Qy 301 LLLLHYSDPSAWGNLYGAVPSLALSTLNSCVDPFYIVYVSAEPRDKVRAGLFORS- 360
Db 307 LLLVHYFLIKTRQSHVYALYVLCGLSTLNSCIDFVYFYFYSKDFRDHARNALLCRSV 366
Qy 360 -PGDTVASKASAGSGRGMTHSS 382
Db 367 RTVARMQITSLSSNFKSRKSGSYSS 390

RESULT 3
US-08-476-976-61
; Sequence 61, Application US/08476976
; Patent No. 5874400
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT W.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,976
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/390,301
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ADLER, REID G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: 2803-0006.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-476-976-61

```

Query Match 26.6% Score 530; DB 2; Length 399;
Best Local Similarity 34.1% Pred. No. 1.3e-34;
Matches 131; Conservative 68; Mismatches 141; Indels 44; Gaps 11;

[illegible]

RECEIVED
US-08-476-000-2
Sequence 2, Application US/08476000
Patent No. 5716789
GENERAL INFORMATION:
APPLICANT: SUNDELIN, JOHAN
APPLICANT: SCARBOROUGH, ROBERT M.
TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/476,000
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/390,301
FILING DATE: 25-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: ADLER, REID G.
REGISTRATION NUMBER: 30,988
REFERENCE/DOCKET NUMBER: 2803-0006.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
TELEX: 90-4030
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 395 amino acids
TYPE: amino acid
TOPOLOGY: linear

Query Match 26.5%; Score 528; DB 1; Length 395;
Best Local Similarity 34.1%; Pred. No. 1.9e-34;

```
; MOLECULE TYPE: protein
US-08-476-000-2

Query Match      26.5%; Score 528; DB 1; Length 395;
Best Local Similarity 34.1%; Pred. No. 1.9e-34;
Matches 131; Conservative 67; Mismatches 142; Indels 44; Gaps 11;

QY 13 GFSLSGGTQT-PSVYDESGTGGDDSTPSILPAPRGYPGVQVCANDSDTLELPDSSRALL 71
Db 33 GRSILGRLETPPI-----TGKG-----VPVEPGF-----SIDEFSASIL 67
QY 72 LGWVPTRLVPALYGLVVLVGLPANGALWV-LATQAPRLPSTMLLMNLATADLLALALP 130
Db 68 TGKLTIVFLPVVYIIIVFVIGLPSNGMALWIFLRTKKHPAVIYMANLADLLSVIWF 127
QY 131 PRIAYHLRGQRPFGGAACRLATAALYGHMYGSVLLLAAYSDRYLALVHPL---RARAL 187
Db 128 LKISYHLHGNWVYGEALCKVLIGFFYGNMYCSILFMTCLSVQRYWVIVNPMGHPKKA- 187
QY 188 RGRRLALGLCMAWLAALALPLTLQRTFLRSDRVLCHDALPLDAQASHWQPAFTC 247
Db 187 ---NIAGVSLAIWLILFVITPLVYMKOTIYPALNITTTCHDVLPVEVLVGMDFNYFLS 243
QY 248 LALLGCF-LPLAMLLCYGATLHTLAASG-----RRYGHALRLTAVVLASAVAFVPSN 300
Db 244 LA-IGVFLPALLTASAYVLMIKTLSSAMDEHSENKQRAIRLIITVLMYFICFAPS 302
QY 301 LLLLHYDPSAMGNLYGAVVPSLALSTLNSCVDPFYIIYVSAEFRKVRAGLFORS- 360
Db 303 LLLVHYFLIKTQQRSHVYALYVALCLSTLNSCIDPFYVYFVSKDFRDHARNALLCRSV 362
QY 360 -PGDTVASKASAEGRSGMGTHSS 382
Db 363 RTVNRMQISLSSNKFSRKGSGSYSS 386

RESULT 6
US-08-472-840-2
; Sequence 2, Application US/08472840
; Patent No. 5763575
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/472,840
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,301
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ADLER, REID G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: 2803-0006.20
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763

; MOLECULE TYPE: protein
US-08-476-000-2

Query Match      26.5%; Score 528; DB 2; Length 395;
Best Local Similarity 34.1%; Pred. No. 1.9e-34;
Matches 131; Conservative 67; Mismatches 142; Indels 44; Gaps 11;

QY 13 GFSLSGGTQT-PSVYDESGTGGDDSTPSILPAPRGYPGVQVCANDSDTLELPDSSRALL 71
Db 33 GRSILGRLETPPI-----TGKG-----VPVEPGF-----SIDEFSASIL 67
QY 72 LGWVPTRLVPALYGLVVLVGLPANGALWV-LATQAPRLPSTMLLMNLATADLLALALP 130
Db 68 TGKLTIVFLPVVYIIIVFVIGLPSNGMALWIFLRTKKHPAVIYMANLADLLSVIWF 127
QY 131 PRIAYHLRGQRPFGGAACRLATAALYGHMYGSVLLLAAYSDRYLALVHPL---RARAL 187
Db 128 LKISYHLHGNWVYGEALCKVLIGFFYGNMYCSILFMTCLSVQRYWVIVNPMGHPKKA- 187
QY 188 RGRRLALGLCMAWLAALALPLTLQRTFLRSDRVLCHDALPLDAQASHWQPAFTC 247
Db 187 ---NIAGVSLAIWLILFVITPLVYMKOTIYPALNITTTCHDVLPVEVLVGMDFNYFLS 243
QY 248 LALLGCF-LPLAMLLCYGATLHTLAASG-----RRYGHALRLTAVVLASAVAFVPSN 300
Db 244 LA-IGVFLPALLTASAYVLMIKTLSSAMDEHSENKQRAIRLIITVLMYFICFAPS 302
QY 301 LLLLHYDPSAMGNLYGAVVPSLALSTLNSCVDPFYIIYVSAEFRKVRAGLFORS- 360
Db 303 LLLVHYFLIKTQQRSHVYALYVALCLSTLNSCIDPFYVYFVSKDFRDHARNALLCRSV 362
QY 360 -PGDTVASKASAEGRSGMGTHSS 382
Db 363 RTVNRMQISLSSNKFSRKGSGSYSS 386

RESULT 7
US-08-476-976-2
; Sequence 2, Application US/08476976
; Patent No. 5874400
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,976
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/390,301
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
```

```

US-08-476-000-3
; Sequence 5, Application US/08476000
; Patent No. 5716789
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,000
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435

```

```
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/390,301
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ADLER, REID G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: 2803-0006.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 395 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-476-000-5

Query Match 26.4%; Score 525; DB 1; Length 395;
Best Local Similarity 34.8%; Pred. No. 3.2e-34;
Matches 125; Conservative 63; Mismatches 129; Indels 42; Gaps 10;

QY 13 GFSLSGGTQT-PSVYDESGTGDDSTPSILPAPRGYQGVCAANDSDTLELPDSSRALL 71
Db 33 GRSLIGRLTQPPi-----TKG-----VPVEPGF-----SIDEFSASIL 67
QY 72 LGWVTRLPALYGLVVLVGLPANGIALWV-LATQAPRLPSTPMLNMLNATADLLALALP 130
Db 68 TGLKTLVFLPVYIIIVFVIGLPSNGMALWIFLRTKKKHPAVIYMANLADLLSVINFP 127
QY 131 PRIAYHLRGQRPFGAACLATAALYGHMYGVSLLAAVSLDRYLALVHPL---RARAL 187
Db 128 LKISYHLGNWVYGCALCKVLIGFYGNMYCSILPTCLSVQRYVWVYVNPNGHPRKKA- 187
QY 188 RGRRLALGCMALWMAAALALPLTLQRTFLARSDRVLCHDALPLDAQASHWQPAFTC 247
Db 187 ---NIAVGSLALWLLIFLVTIPLYVMKQTIYPALNITTCDDVLPVEVLGDMENYFLS 243
QY 248 LALGCGFL-PLLAMLLCYGATLHTLAAG-----RRYGHALRLTAVVLASAVAFVPSN 300
Db 244 LA-IGVFLFPALLTASAYVLMIKTLRSSAMDEHSEKKRQRAIRLIITVLAMYFICFAPS 302
QY 301 LLLLHYSDPSPSANGNLYGAYVPSLALSTLNSCDVPFIYVYSAEFRDKVRAGLFQRS 359
Db 303 LLLVHYFLIKTQRQSHVYALYLVALCLSTLNSCIDPFYVYFVSKDFRDHARNALLCRS 361

RESULT 10
US-08-472-840-5
; Sequence 5, Application US/08472840
; Patent No. 5763575
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,840
```

```
;
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,301
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ADLER, REID G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: 2803-0006.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 395 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-472-840-5

Query Match 26.4%; Score 525; DB 2; Length 395;
Best Local Similarity 34.8%; Pred. No. 3.2e-34;
Matches 125; Conservative 63; Mismatches 129; Indels 42; Gaps 10;

QY 13 GFSLSGGTQT-PSVYDESGTGDDSTPSILPAPRGYQGVCAANDSDTLELPDSSRALL 71
Db 33 GRSLIGRLTQPPi-----TKG-----VPVEPGF-----SIDEFSASIL 67
QY 72 LGWVTRLPALYGLVVLVGLPANGIALWV-LATQAPRLPSTPMLNMLNATADLLALALP 130
Db 68 TGLKTLVFLPVYIIIVFVIGLPSNGMALWIFLRTKKKHPAVIYMANLADLLSVINFP 127
QY 131 PRIAYHLRGQRPFGAACLATAALYGHMYGVSLLAAVSLDRYLALVHPL---RARAL 187
Db 128 LKISYHLGNWVYGCALCKVLIGFYGNMYCSILPTCLSVQRYVWVYVNPNGHPRKKA- 187
QY 188 RGRRLALGCMALWMAAALALPLTLQRTFLARSDRVLCHDALPLDAQASHWQPAFTC 247
Db 187 ---NIAVGSLALWLLIFLVTIPLYVMKQTIYPALNITTCDDVLPVEVLGDMENYFLS 243
QY 248 LALGCGFL-PLLAMLLCYGATLHTLAAG-----RRYGHALRLTAVVLASAVAFVPSN 300
Db 244 LA-IGVFLFPALLTASAYVLMIKTLRSSAMDEHSEKKRQRAIRLIITVLAMYFICFAPS 302
QY 301 LLLLHYSDPSPSANGNLYGAYVPSLALSTLNSCDVPFIYVYSAEFRDKVRAGLFQRS 359
Db 303 LLLVHYFLIKTQRQSHVYALYLVALCLSTLNSCIDPFYVYFVSKDFRDHARNALLCRS 361

RESULT 11
US-08-476-976-5
; Sequence 5, Application US/08476976
; Patent No. 5874400
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```


;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/476,976
;; FILING DATE: 07-JUN-1995
;; CLASSIFICATION: 536
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/390,301
;; FILING DATE: 25-JAN-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: ADLER, REID G.
;; REGISTRATION NUMBER: 30,988
;; REFERENCE/DOCKET NUMBER: 2803-0006.20
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 887-1500
;; TELEFAX: (202) 887-0763
;; TELEX: 90-4030
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 395 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-476-976-5

Query Match 26.4%; Score 525; DB 2; Length 395;
Best Local Similarity 34.8%; Pred. No. 3.2e-34;
Matches 125; Conservative 63; Mismatches 129; Indels 42; Gaps 10;

QY 13 GFSLSGGTQ-PSYVDESGTGGDDSTPSILPAPRGVPGVCANDSTLELPDSSRALP 71
DB 33 GRSLIGLETPPI-----TGKG-----VPVEPGE-----SIDFSAIIL 67
QY 72 LGWVPTRLVPALYGLVVLGVPANGALWV-LATQAPRLPSTMLMNLATADLLALALP 130
DB 68 TGKLTTFVLPVYIIVFVIGLPSMGALWIEFTKPKHPAIVYMANLALADLLSVIWF 127
QY 131 PRIAYHLRGQWPGEACRATATAALYGHMYGVSLLLAANVSDRYLALVHPL---RARAL 187
DB 128 LKISYHLGNWVYGEALCKVLIGFFYGNMYCSILFTCLSVQRYVIVNPMGHPKKA- 187
QY 188 RGRRLALGLCMAAALALPLTLQRTPLARSQVRLCHDALPLDAQASHWQPAFTC 247
DB 187 ---NIAVGSIAWLLFLVPIVYVVKQTIYIPALNITTCVDVLPVEVLGDMFNELS 243
QY 248 LALGCEL-PLLAMLLCYGATLHTLAAG-----RRYGHALRLTAVVLASAVAFFVPSN 300
DB 244 LA-IGVLEFPALLTASAYVLMIKTLRSSAMDEHSEKKRQRAIRLIITVLMYFICFAPS 302
QY 301 LLLLLHYSDPSPSANGNLYGAYVPSLALSTLNSCVDPFIYVYVSAERDKYRACLQFS 359
DB 303 LLLVVHYFLIKTRQSHVYALYVALCLSTLNSCIDPFYVYVSKDFRDHARNALLCRS 361

RESULT 12
US-08-742-440A-6
; Sequence 6, Application US/08742440A
; Patent No. 5892014
; GENERAL INFORMATION:
; APPLICANT: Coughlin, Shaun
; APPLICANT: Ishihari, Hiroaki
; APPLICANT: Connolly, Andrew
; TITLE OF INVENTION: Protease Activated Receptor
; NUMBER OF SEQUENCES: 3 and Uses Thereof
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Avenue, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSeq for Windows Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/742,440A
;; FILING DATE: 30-OCT-1996
;; CLASSIFICATION: 536
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Sherwood, Pamela J
;; REGISTRATION NUMBER: 36,677
;; REFERENCE/DOCKET NUMBER: UCAL/060PAT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 650-327-3400
;; TELEFAX: 650 327-3231
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 408 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; FEATURE:
;; US-08-742-440A-6

Query Match 26.2%; Score 521.5; DB 2; Length 408;
Best Local Similarity 33.8%; Pred. No. 6.2e-34;
Matches 120; Conservative 62; Mismatches 158; Indels 15; Gaps 5;
QY 3 GRLLMLVLGFLSGGTQ-----TPSVYDESGTGGDDSTPSILPAPRGVQV---- 54
DB 29 GLLLLLTFQCSGMENDNNAKPTLPKTFRGAPPNSFEFP--FSALEGWGTATIVK 86
QY 54 --CANDSDT-LELPDSSRALLLGWVPTRLVPALYGLVVLGVPANGALWVLAQAPRLP 110
DB 87 IKPEESASLHVKNATMGYLTSLSTKLIPAIYLLVGVVGPANAVTLMWLFRTSIC 146
QY 111 STMLMLNATADLLALPPIAYHLRGQWPGEACRATATAALYGHMYGVSLLLAALV 170
DB 147 TTYFTYNTALADLEFCVTLFPFKYAYHLGNWVGEVLCRATTVIFGNYCSILLACI 206
QY 171 SLDRYLAHVPLRARALRGRLALGLCMAAALALPLTLQRTPLARSQVRLCHD 230
DB 207 SINRYLAIVHPVTYRGVLPKHTYALVTCGLVWATVFLYMLPFFILKQEVYLVQPDIT 266
QY 231 ALPDAQASHWQ-PAFTCLALLGCFLLPAMLLCYGATLHTLAASGRYGHALRLTAVL 289
DB 267 VHNTESSSPFQLYYFISLAFPGFLIPFVLIYCYAAIIRLNAYDHRWLMYVKASLL 326
QY 290 ASAVAFVPSNLLLLHYSDPSPSANGNLYGAYVPSLALSTLNSCVDPFIYVYV 344
DB 327 VIFTICFAPSNIILIIHANYNNNTDGLFYIILALCLGSLNSCLDPFLYFLMS 381

RESULT 13
US-08-097-938-6
; Sequence 6, Application US/08097938
; Patent No. 5623174
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR AND ITS
; AGONISTS AND ANTAGONISTS
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20006-1812

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/097,938
; FILING DATE: 26-JUL-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 22803-20006.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 398 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-097-938-6

```

Query Match	25.6%	Score	510;	DB	1;	Length	398;
Best Local Similarity	36.5%	Pred.	No. 4.9e-33;				
Matches	110;	Conservative	58;	Mismatches	123;	Indels	10;
Gaps							
Qy	67	SRALLGQVTRLPALYGLVVLVGLPANGALAV-LATQAPRLPSPYMLNLTADALLL	125				
Db	66	SASVLTGKLTVPUPYIIIVFVGLPSNGMALVFLFRTKKKPAVIYMANLALADLS	125				
Qy	126	ALALPPRIAYHLRGORWPFGEAACRLATAALYGHMYSVLLAAVSLDRYLALVHPLRAR	185				
Db	126	VINEPLAIYHIHGNWYIGALCNVLGFFYGNMYSILEFTCLSVQRYVVIYNPM-GH	184				
Qy	186	ALGRRLALGCMAMLMALALPTLRQTRFLRSDRVLCHDPLDQAQSHWQPAF	245				
Db	185	SRKNAIAIGISLAIMLLILVITPLVYVVKOTIEIPALNITTHDVLPEQVLVGDMFNYF	244				
Qy	246	TCIALLCCEL-PLLAMLLCYATHTLAASG-----RRYGHARLRLTAVVLASAVAFVP	298				
Db	245	LSLA-IGVFLPAPLTASVYVLMIRLSSAMDENSEKKRRAIKLIVTLYAMLYCIFP	303				
Qy	299	SNILLLLGHSDPSAGNLYGAVVPSLALSTLNSCDVDPFFIYVSVAFDRKVRAGLFR	358				
Db	304	SNLLVVHYFLIKSQGSHVYALIVALCLSTLNSCIDPFFYVYVSHDFDRHAKNALLCR	363				
Qy	359	S	359				
Db	364	S	364				

```

RESULT 14
US-08-476-000-6
; Sequence 6, Application US/08476000
; Patent No. 5716789
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,000
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/390,301
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ADLER, REID G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: 2803-0006.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 398 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-476-000-6

Query Match          25.6%; Score 510; DB 1; Length 398;
Best Local Similarity 36.5%; Pred. No. 4.9e-33;
Matches 110; Conservative 58; Mismatches 123; Indels 10; Gaps 5;

QY   67 SRALLIGWVTRLPALYGVLVGLPANGIALWV-LATQAPRLPSTMLMNLATADLLL 125
      |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    66 SASVLTGKLTVPPIVIIVVVGLPSNGMALWVFRTFKKHHPAVIYMANLADLLS 125

QY   126 ALAPPRIRAYHLRGQRPFGEACRLATAALYGHMYGSVLLLAASLDYLALVHPLRAR 185
      |:|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||:
Db    126 VINFPKIAYHIHGNNIYGEALCNVLGIFFYGNMYCSILFTCLSVORYWVIVNPN-GH 184

QY   186 ALGRRLAIGLCHAAWLMAALALPLTLRQTFLARSDRVLCHDALPLDAQASHWQPAF 245
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    185 SRKANTAIGSLAIWILLITLVPLVVKQTIFIPALNTTCHDYLPEQVLVGDMFNIF 244

QY   246 TCLAALLCGFL-PLAMLLCYGATLHTLAAGS-----RRYGHALRLTAVVLASAVAFVP 298
      ||:|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    245 LSLA-IGVFLPFAFLTASAVLMRLRSSAMDENSEKKRKRAIKLIIVTLAWYLICFIP 303

QY   299 SNLLLLHYSDSPSGAWNLYGYVPSLSALTNSCVDPPFIYYVVSAEFRDKVRAGLFQR 358
      ||||::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    304 SNLLVVHYFLIKSQGSHVYALIVYALCLSTLNCLSDPFIYFYVSHDFDRHARNALLCR 363

QY   359 S 359
Db    364 S 364

RESULT 15
US-08-472-840-6
; Sequence 6, Application US/08472840
; Patent No. 5763575
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA

```

```

; ZIP: 20006-1812
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC Compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/08/472,840
;
; FILING DATE:
;
; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER: US/08/390,301
;
; FILING DATE: 25-JAN-1995
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: ADLER, REID G.
;
; REGISTRATION NUMBER: 30,988
;
; REFERENCE/DOCKET NUMBER: 2803-0006.20
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (202) 887-1500
;
; TELEFAX: (202) 887-0763
;
; TELEX: 90-4030
;
; INFORMATION FOR SEQ ID NO: 6:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 398 amino acids
;
; TYPE: amino acid
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
; US-08-472,840-6

```

Query Match	25.68;	Score	510;	DB	2;	Length	398;
Best Local Similarity	36.5%;	Pred.	No. 4.9e-33;				
Matches	110;	Conservative	58;	Mismatches	123;	Indels	10;
Gaps							
QY	67	SRALLGQVTRLPALYGLVVLVGLPANGIALAV-LATQAPRLPSTMLLNLTATADLL	125				
DB	66	SASVLTGKLTVFUPTIIVVVLVGLPSGMLWVLFRTKKKHPAVIYMANLALADLLS	125				
QY	126	ALALPPRIAYHLRGQRPFGGEAACRLATAALYGHMYGSVLLAAVSLDRYALVHPLRAR	185				
DB	126	VWFEPLAIYHIHGNWYGEALCNVLIGFPYGNMYCSILFMTCLSVQRYVWVYNPM-GH	184				
QY	186	ALGRRLALGCMAMLMMAALPLTLQRTFLRSLDRVLCHDAPLPLDAQASHWQAPAF	245				
DB	185	SRKKAITAIGSLAIWLLILVITPLYVVKOTIEIPALNITTTCHDVLPEQVLVGDMENYF	244				
QY	246	TCIALLCCEL-PLIAMLCCYATLHTLAASG-----RRYCHALRLTAVVLASAVAFVP	298				
DB	245	LSLA-IGVFLFPALFTASVYVLMIRLSSAMDENSEKKRRAKRLIYTVLAMYLTICIFP	303				
QY	299	SNLILLHRYDPSPSANGNLYGAVPSIALSTLNSCVDPPFYIYVVSABFDRKVRAGLGF	358				
DB	304	SNLLVVHYFLIKSQGSHVYALVIALCLSTLNSCIDPFYIFYVSHDFDRHAKNALLCR	363				
QY	359	S	359				
DB	364	S	364				

Search completed: October 31, 1999, 06:15:14
Job time: 2194 sec

This Page Blank (uspio)

Db	860	TGCTTGAGGCCAGCAGTTCAAGACCAGCCTGGGCAACATAGCAAGACCCCTGTCTTTATTT	919
QY	1640	AAAAACAAAAA-----TTAGCTGGGCTTGCTGGCTGGGCGCTGTGA	1680
Db	920	AAACCAAAAAAAAAAAGACAAGCTTAGCCAGCGCATGGTGGCAGTTGCGTGTGA	979
QY	1681	ATCCAGCTACTCAGAGACTGAGGCAGAAAGATCG-CTTGAACCTGGGAGCGAGAGTT	1739
Db	980	GTCCAGGTACTCAGGAGCTGAGATAGAAGGATGTCTTGAGCCCAAGAAATCAAGGCT	1039
QY	1740	GCAGTGAGCCGAGATTGCGCCACTGACTCCAGCCTCGGTGACAGAGAGCCTGTCTCTAA	1799
Db	1040	GTAGTGAGCTGATGTGTACCACTCAGCTCCAGCCTGGGTGACAAAG-----	1087
QY	1800	ATTAATTAAATTAATTAATTAATCAATTTTAAAAAGACGAAAAAGTCACGCCAGGTGCA	1859
Db	1087	-----CAAAACACTGTCTCCAAAAAATTTAGGCTTTGGCAAGCGCA	1129
QY	1860	GTGGCTCAGCCCTATAATCTCAGCACTCTGGGAGGCC-AGATGGAGGATTTGTTGAAGC	1918
Db	1130	CGGCTCAGCCCTGTGTCCAGCACTTTGGGAAGCCGAAGCAGCGAGATCACTTGAGGT	1189
QY	1919	CAGGAGTTTGGGACAGCCTGGGCAACATAGGGGATCCCATCTCTACACACAAAAAAT	1978
Db	1190	CAGGAGTTGGAGACAGCCTGGCCAACTGTGAACCTGTCTCTACTGAAAATACAA	1249
QY	1979	TTTTTAATCAACCAGGCATTGTGGCATCGCCTATAGTCCCAGCCACTCAAGAGGCACAG	2038
Db	1250	AATTACCCGGTTGGTAGTGCGTCTGG-----TAATCTTAGCTACTTGGGAGGTGAG	1305
QY	2039	CGGGGAGGA--TCACTTGAGCCTGGGAGGTTGTGGTGCAGTGAAGCTATGATTGTACCAC	2096
Db	1306	GCAGGGGGAATTGCCCTGAAACCTGGAGCGGAGGCTGCAGTGAGCCGAGATTGCATCAT	1365
QY	2097	TGCATCTAGCCTGGGCAACAGAGCAAGACCTTGTCTCAAAATPAACAACATAAATTA	2156
Db	1366	TGCATCTTAGCTTGGACACAGAGCTAGACTCCATCCCAAAAAAAAAAAAAA-----	1420
QY	2157	AAAAAGAGACGAGAGATAGTGGGTGTGGTGTCAACCTGCAACTCGCAATCCAGCAGCTTGG	2216
Db	1420	-----AGTAGCGGGCAGGTTGGCTCAGCCGTGTAAATCCAGCAGCTTTGG	1464
QY	2217	AAGCCGAGGTGGGAGATCATCTGAGGCCAGGAGTTCAAGACAGCAGCTGGCTAAACATGG	2276
Db	1465	GAGGCCGAGCGGCGGATCA--TGAGGGCAGGAGATCGAGCACCCTCTGGCTAACACGG	1522
QY	2277	TGAATCTCATCTTACCAAAAAATAC-AAAAATTAGCCAGCGGTGGTGGGGCACCTGT	2335
Db	1523	TGAACCCCTGTCTCTACTAAAAAATACAAAAAATTAGCCGCGAGGTGGCGGGCGCTGT	1582
QY	2336	ACTGGGAGGTGCCACCAGCTACTGGGAGGCTGAGTCAGGAGAAATCCCTTGAACCTG	2395
Db	1583	AGT-----CCAGCTACTCAGGAGAGTGAGCCAGGAGAAATGGCTGTGAACCCG	1629
QY	2396	GGAGCGGAGGTTGGGTCAGCTGAGATGGTGCACCTGCATCTCCAGCCTGGGCGAAAGAG	2455
Db	1630	GGGGCGGAGCTGCAGTGAGCCGAGATCGGCCACTGCATCTCAGCTTGGGTGACACCG	1689
QY	2456	CGACTCTGTCTCCAAAAAAGAGAA	2481
Db	1690	AGACTCGCTCTCAAAAAAATAAATA	1715

RESULT

RESULT 3
US-08-484-044-10
: Sequence No. 10, Application US/08484044
: Patent No. 5552282
: GENERAL INFORMATION:
: APPLICANT: Caskey, C. T.
: APPLICANT: Fu, Ying-Hui
: APPLICANT: Friedman, David L.
: APPLICANT: Pizzuti, Antonio
: APPLICANT: Fenwick, Raymond G.

Db 6757 AGGAGTTCGAGACCCAGCCTGACCAAGCTGGAGAAACCTCATCTCTACTAAAAACAA--- 6814
Qy 1980 TTTTAATGAACAGGAGCATTTGGCATGCGCTATAGTCCAGCCTCAAGAGGCACAGG 2039
Db 6814 ----AATTAGCAGGAGCATGGTGCAATACCTGTATCCAGCTACTCAGGAGGCTGAGG 6869
Qy 2040 CGGGAGGATCACTTGGAGCTGGGAGGTTGGTTGCAGTGAGCTATGATTGTACCACTGC 2099
Db 6870 CAGGAGAATCA---GAACCCAGGAGGGAGGTTGGTTAGCTGAGATCGTGCCATTGC 6926
Qy 2100 ACTCCAGCCTGGGCAACAGCAGACAGACCTTGTCTCAAAAATTAACAAACTAAAAATAAA 2159
Db 6927 ATTCCAGCCTGGGCAACAGAGATGAACCTTCATCTCAAAAAAATAAA 6978
Qy 2160 AAAGAAGACGAGAGATAGTGGGTGGTGGCTCACACCTGCAATCCAGCAGCTTTGGAG 2219
Db 6979 AGTACTATTATACAGGCTGGGATGGTGGCTCACGCTTGGAAATCCAGCAGCTTTGGGAG 7038
Qy 2220 GCGGAGGTGGGAGATCATCTGAGGCCAGGAGTTCAAGACCAGCCTGGCTAACTGGTGA 2279
Db 7039 GCTGAAGTGGAGGATGCTTCAGGCCAGGAGTTCAAGACCAGCCTGAGCAACATAATGA 7098
Qy 2280 AATCTATCTCTACCAAAAATAACAAAAT---AGCCAGGCGTGGTGGGACCTGTAC 2337
Db 7099 GACCCGTGCTCTACAAAAAATTAACAAAATTCGTCAGCAGCATGGTGGTCTGTCGCTGCAG 7158
Qy 2338 TGGGGAGGTGCCACCCAGCTACTGGGAGGCTGAGTCAAGGAGATCGCTTGAACCTGG 2396
Db 7159 T-----CCTAGCTACTAGGAGTCTGAAGTAGGAGATCACTTGGAGCTGG 7204

RESULT 4

US-08-080-255-6
: Sequence 6, Application US/08080255
: Patent No. 5487970
: GENERAL INFORMATION:
: APPLICANT: Rowley, Janet D.
: APPLICANT: Diaz, Manuel O.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
: TITLE OF INVENTION: DEPOSITING GENE REARRANGEMENTS AND
: TITLE OF INVENTION: TRANSLOCATIONS
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESS: Arnold, White & Durkee
: STREET: P. O. Box 4433
: CITY: Houston
: STATE: Texas
: COUNTRY: USA
: ZIP: 77210
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/080,255
: FILING DATE: 19930617
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Parker, David L.
: REGISTRATION NUMBER: 32,165
: REFERENCE/DOCKET NUMBER: ARCD:072/PAR
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (512) 320-7200
: TELEFAX: (512) 474-7577
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 8392 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-080-255-6

Query Match 6.68; Score 324.6; DB 1; Length 8392;
Best Local Similarity 66.9%; Pred. No. 1.6e-50;
Matches 658; Conservative 0; Mismatches 229; Indels 96; Gaps 10;

Qy 1520 GCAGGCGCTGGTGGCTCACGCTGTAAATCCAGCAGCTTTAAGAGGCCCAAGCGGATGGAT 1579
Db 860 GCTGGGACGGTGGCTCACGCTGGTAAATCCCAACACTTAGTGAGGCTGAGGTGGGAGAT 919
Qy 1580 CACTTGAAGCCAGGAGTTCAACACCCAGCCTGAGCAACATGTTAAACCCCTCTCTACCA 1639
Db 920 TGCITGAGGCCAGCAGTTCGAAGACGACGCTGGGCAACATAGCAAGACCCCTCTCTTATT 979
Qy 1640 AAAATACAAAAA-----TTAGCTGGGCTTGGTGGCTGGGCGGCTGTGTA 1680
Db 980 AAACAAAAAATAAAGAAAGAAAGAAAGTATAGCCAGGCAATGGTGGCAGTTGCGTGTGA 1039
Qy 1681 ATCCAGCTACTCAGGAGCTGAGGACAGAGGATCG-CTTGAACCTGGGAGGCGAGAGGTT 1739
Db 1040 GTCCAGGCTACTCAGGAGGCTGAGATAGAGGATTTCTTGAGCCAGGAAATTCAGGCT 1099
Qy 1740 GCAGTGAAGCCAGGATTTGGCCACCTGGAGCTCCAGCTGGGTCACAGAGAGCCCTGTCTAA 1799
Db 1100 GTAGTGAGCTATGATTGTACCACTGCAGTCCAGGCTGGGTGACAAAG----- 1147
Qy 1800 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1859
Db 1147 -----CAAAACACTGTCTCCAAAAAATTTAGGCTTGGCAAGGGCGCA 1189
Qy 1860 GTGGCTACGCTATATATCTCAGCAGCTCTGGGAGGCC-AAGATGGAGGATTTGCTTGAAGC 1918
Db 1190 -CGGCTACGCTGTGATCCAGCAGCTTTGGGAAGCCGAACGAGCAGATCAGTTGAGT 1248
Qy 1919 CAGGAGTTGGGACAGCCTGGGCAACATAGGGGATCCCAATCTCTACACACAAAAAAT 1978
Db 1249 CAGGAGTTGGAGACCGCCTGGCCAAACATGGTGAACCCCTGTCTACTGAAAAATACAA- 1308
Qy 1979 TTTTAAATGAACAGGCAATTTGTCATGCGCTATAGTCCAGCAGCTCAAGAGGCACAG 2038
Db 1308 ----AAATTAGCCGGTTGGTAGTGGTGTGTTGTAATCTTAGTACTTTGGGAGGCTGAG 1363
Qy 2039 GCGGAGGATCACTTGAAGCTGGGAGGTTGGTTGTCAGTGAGCTATGATTGTACCACTG 2098
Db 1364 GCAGGGAATTCCTGAACCTGCGAGCGGAGGCTGCGAGTCAGCGGAGATTGCATCATTTG 1423
Qy 2099 CACTCCAGCCTGGGCAACAGAGACCTTGTCTCAAAAAATAACAAACTAAAAATTA 2158
Db 1424 CACTTAGCTGGACACAGAGCTAGACTCCATCCCAAAAAA----- 1472
Qy 2159 AAAGAAGACGAGAGATAGTGGGTGGTGGCTCACACCTCAATCCAGCAGCTTTGGAA 2218
Db 1472 -----AAAGTAGCCGGCAGCTGGCTCAGGCTGTAAATCCAGCAGCTTTGGGA 1520
Qy 2219 GCGGAGGTGGGAGATCATCTGAGGCCAGGAGTTCAAGACCAGCCTGGCTTAACATGGT 2278
Db 1521 GCGGAGGCGGGCGGATCA--TGAGGCGAGGAGATCGAGACCATCTCTGGCTTAACACGCTG 1578
Qy 2279 AATCTCTATCTTACCAAAAAATAAATAATAGCCAGGCTGGTGGGCGGCTGTACT 2338
Db 1579 AATCCCTGTCTCTATAAATAACAAAAATTAGCCCGGAGGCTGGGCGGCTGTAGT 1638
Qy 2339 GGGGAGGTGCCACCCAGCTACTGGGAGGCTGAGTCAGGAGAAATCGTTGAACCTGGGA 2398
Db 1639 -----CCAGCTACTCAGGAGACTGAGGAGGAGAAATGGGTGAACCCGGG- 1695
Qy 2399 GCGGAGGTTGCGGTGAGTGTGGCAGTGCACCTCCAGCTGGGCGAAAGAGCGA 2458
Db 1685 GCGGAGCCTGCACTGAGCCGAGATCGGCCACTGCACCTCCAGCTGGGTGACACCGAGA 1744
Qy 2459 CTCTGTCTCCAAAAAAGAGAA 2481
Db 1745 CTCCGTCTCAAAAAAATAAATA 1767


```

RESULT 5
PCT-US93-05857-6
; Sequence 6, Application PC/TUS9305857
; GENERAL INFORMATION:
; APPLICANT: Board of Regents
; TITLE OF INVENTION: The University of Texas System
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
; TITLE OF INVENTION: GENE REARRANGEMENTS AND TRANSLOCATIONS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/05857
; FILING DATE: 19930617
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/900,689
; FILING DATE: 17/06/92
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:072/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 320-7200
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8392 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US93-05857-6

Query Match 6.6%; Score 324.6; DB 5; Length 8392;
Best Local Similarity 66.9%; Pred. No. 1.6e-50;
Matches 658; Conservative 0; Mismatches 229; Indels 96; Gaps 10;

QY 1520 GCCAGGCTGTGGCTCACGCTGTATCCAGCAGCTTTAAGAGCGCAAGCGGATGGAT 1579
Db 1520 GCTGGCAGGCTGTGGCTCACGCTGTATCCAGCAGCTTTAAGAGCGCAAGCGGAT 919
QY 1580 CAGTTGAGCCAGGAGTTCAACACCGCTGAGCAACATGTTAAACCCCATCTCTACCA 1639
Db 1580 CAGTTGAGCCAGGAGTTCAACACCGCTGAGCAACATGTTAAACCCCATCTCTACCA 979
QY 1640 AAAATACAAAAA-----TTAGTGGCTGTGGCTGGCGCCCTGTA 1680
Db 1640 AAAATACAAAAA-----TTAGTGGCTGTGGCTGGCGCCCTGTA 1039
QY 1681 ATCCAGCTACTCAGGAGCTGAGCGAAGAGATCG-CTTGAACCTGGGAGCGAGAGTT 1739
Db 1681 ATCCAGCTACTCAGGAGCTGAGCGAAGAGATCG-CTTGAACCTGGGAGCGAGAGTT 1099
QY 1740 GCACTGAGCCAGATGTCGGCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1799
Db 1740 GCACTGAGCCAGATGTCGGCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1147
QY 1800 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1859
Db 1800 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA

```

```

Db 1147 -----CAAAACACTGTCTCCAAAAAATTTAGCTTGGCAAGCGCA 1189
QY 1860 GTGGCTCAGCGCTATATCTCAGCAGCTCTGGGAGGCC-AAGATGGAGGATGCTTTGAAGC 1918
Db 1190 -CGGCTCAGCGCTGTGATCCAGCAGCTTTGGGAAGCGGAGCAGGATGCTTGGAGCT 1248
QY 1919 CAGGAGTTTGGGACCGCTGGGCAACATAGGGGATCCCATCTCTACACACAAAAAAT 1978
Db 1249 CAGGAGTTTGGGACCGCTGGGCAACATAGGGGATCCCATCTCTACACACAAAAAAT 1308
QY 1979 TTTTATGAACCCAGGCTTTGGCATGCGCTATAGTCCAGCCACTCAAGAGGACAG 2038
Db 1308 -----AAATTAGCGGTTGGTAGTGGGTGCTTGTATCTAGCTAGCTTGGGAGGCTGAG 1363
QY 2039 GCGGAGGATCAGCTTGGAGCTGGGAGGTTGTGGTTGAGCTGAGCTATGATGACCACTG 2098
Db 1364 GCAGGGGAATTGCTGCACTGCGAGGCGGAGGCTGAGTGGAGGATGCTGATCATTG 1423
QY 2099 CACTCCAGCTGGGCAACAGAGCAGACCTTGTCTCAAAAAAATAAACAATAAATAA 2158
Db 1424 CACTTAGCTGGCAACAGAGCTAGACTCCATCCAAAAAATAAATAAATAAATAA 1472
QY 2159 AAAAGAGAGCAGAGATAGTGGGTGGTGGCTCACACCTGCAATCCAGCAGCTTTGGAA 2218
Db 1472 -----AAAGTAGCGGCGGAGTGGTGGTGGCTCACACCTGCAATCCAGCAGCTTTGGAA 1520
QY 2219 GCGGAGGTTGGCAGATCATCTGAGCCAGGAGTTTCAAGACCGAGCTGCTTAACATGGTG 2278
Db 1521 GCGGAGGTTGGCAGATCATCTGAGCCAGGAGTTTCAAGACCGAGCTGCTTAACATGGTG 1578
QY 2279 AATCTCTCTACCAAAAAATAGCCAGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2338
Db 1579 AATCTCTCTACCAAAAAATAGCCAGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1638
QY 2339 GGGGAGGTTGCCACCGAGCTGAGTGGGAGGCTGAGTGGGAGGAGTGGTGGGAGGAG 2398
Db 1639 -----CCAGCTACTCAGGAGTGGGAGGAGTGGTGGGAGGAGTGGTGGGAGGAG 1685
QY 2399 GCGGAGGTTGGGTCAGCTGAGATGGTGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 2458
Db 1685 GCGGAGGTTGGGTCAGCTGAGATGGTGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 1744
QY 2459 CTCTGTCTCAAAAAAAGAGAA 2481
Db 1745 CTCTGTCTCAAAAAAAGAGAA 1767

```

```

RESULT 6
PCT-US95-07201-43/C
; Sequence 43, Application PC/TUS9507201
; GENERAL INFORMATION:
; APPLICANT: Chader, Gerald J.; Becerra, Sofia
; APPLICANT: Patricia; Schwartz, Joan P.;
; APPLICANT: Tanivaki, Takayuki
; TITLE OF INVENTION: PIGMENT EPITHELIUM
; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
; TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07201
; FILING DATE: 06-JUN-1995

```


CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-4530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/273,411
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: WUI06
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 815-6508
TELEFAX: (404) 815-6555
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3373 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc_feature
LOCATION: 104..1201
OTHER INFORMATION: /note= "Nucleotides 104 thr
PUBLICATION INFORMATION:
AUTHORS: Larsen, et al.
JOURNAL: Proc. Nat'l Acad. Sci. USA
VOLUME: 87
PAGES: 6674-6678
DATE: 1990
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 3
US-08-273-411-2

[illegible]

Db	2024	ATCCAGCACATTTGGGAGGCCAAGACACAGGTGGATGACA - AGTCAGGAGATCGAGACCAT	1966
Qy	1936	CCTGGCAACATAGGGGGATCCCATCTCTACACAAAAAAATTTTTTAATGACCAAGGC	1995
Db	1965	CCGGCCAAACATGGTGAACCCCTGCTCTAC ----TAAAAATACAAAAATTAGCCAGGC	1911
Qy	1996	ATTGTGCGATCGCGCTATATGTCGAGCCATCAAGAGGCACAGCGGGAGGATCACTTGA	2055
Db	1910	ATAATGGCCTGTGCTGTAGTCCCATCTACTCAGGAGGCTGAGACAGGAGAATCGCTTGA	1851
Qy	2056	GCCTGGGAGGTGTGGTTGCAGTGCAGCTATGTTACCACCTGCACCTCCAGCCTGGGCAA	2115
Db	1850	ACACAGAGGCGGAGTTGCAGTGTGCGAAGATCACGCCACTGCACCTCCAGCCTGGGCAA	1791
Qy	2116	CAGACGAGACCTTGTCTCAAAAATTAACAACAACTAAATTTAAAAAAA	2162
Db	1790	CAGAGCGAGACCCCTGTCTCCAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA	1744

```

RESULT 10
US-08-525-058A-5/c
; Sequence 5, Application US/08525058A
; Patent No. 5770420
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,058A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lavalleye, Jean-Paul M. P.
; REGISTRATION NUMBER: 31,451
; REFERENCE/DOCKET NUMBER: 2363-060-55
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8174 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
US-08-525-058A-5

```

	Query Match	6.3%	Score 306.6;	DB 3;	Length 8174;
	Best Local Similarity	74.7%;	Pred. No. 3e-47;		
	Matches 483;	Conservative 0;	Mismatches 144;	Indels 20;	Gaps 7;
Qy	1519	GGCCAGGCGCTGGTGGCTCAGCGCTGTATCCACGACCTTTTAAGAGGCCCAAGCGCGATGGA	1578		
Db	6955	GGCCAGGCGATGGTGGCTCAGCGCGGTATCCACGACCTGTGGGAGGCCGAGGTGGGCAGA	6896		

Sequence 3, Application PC/TUS9100899

GENERAL INFORMATION:
APPLICANT: Lowe, John B.
TITLE OF INVENTION: Method and Products For the Synthesis of
TITLE OF INVENTION: Oligosaccharide Structures on Glycoproteins, Glycolipids,
TITLE OF INVENTION: or as Free Molecules, and For the Isolation of Cloned
TITLE OF INVENTION: Genetic Sequences That Determine These Structures
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/00899
FILING DATE: 19910214
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye Ph.D., Jean-Paul
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-021-55 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-5940
TELEFAX: (703)486-2347
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 8174 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Blood
FEATURE:
NAME/KEY: misc.feature
LOCATION: 4686..5780
OTHER INFORMATION: /label= mat_peptide
PCT-US91-00899-3

Query Match 6.3%; Score 306.6; DB 5; Length 8174;
Best Local Similarity 74.7%; Pred. No. 3e-47;
Matches 483; Conservative 0; Mismatches 144; Indels 20; Gaps 7;

QY 1519 GCCACGGCTGGTGGCTCAGCGCTGTATCCAGCAGCTTTAAGAGCCCAAGCGCGATGGA 1578
DB 6955 GCCACGGCTGGTGGCTCAGCGCTGTATCCAGCAGCTTTAAGAGCCCAAGCGCGATGGA 6896
QY 1579 TCACCTTGAGCCGAGGTTCAACACGAGCTGAGCAACATGCTAAACCCCATCTCTACC 1638
DB 6895 TCAC--GAGCTGAGGATCGAGACCATCTGTTAAACACGATGAACCCCGCTCTACT 6838
QY 1639 AAAAATAC---AAAAATAGCTGGGCTGGTGGCTGGGCGCTGTATCCAGCTACTCAG 1695
DB 6837 AAAAATACAAAAAATAGCCGGCGTGGTGGGCGCGCTGTAGTCCAGCTACTCAG 6778
QY 1696 GAGACTGAGGAGAGGATCGCTTGAACCTGGGAGGAGAGGTTGAGTGAGCCGAGATT 1755
DB 6777 GAGGCTGAGGAGAGGATCGCTTGAACCTGGGAGGAGAGGTTGAGTGAGCCGAGATT 6721
QY 1756 GCGCCACTGGAGCTCCAGGCTGGTGGAGAGAGAGCTGTCTCTAAATTAATTAATTA 1815
DB 6720 GCGCCACTGGAGCTCCAGGCTGGTGGAGAGAGAGCTGTCTCTAAATTAATTAATTA 6663

QY 1816 ATTTAATTCATTTTAAAGACGAAAGAGTGCAGCCAGGTGCAGTGGCTCAGCGCTATA 1875
DB 6662 ATAAATAAATAATTAATAA---AAATATCCGGGCTGGGACAGTGGCTCATGCCAGTA 6607
QY 1876 ATCTCAGCAGCTCTGGGAGCCCAAGATGGAGGATGCTTTGAAGCCAGAGATTGGGACCCAG 1935
DB 6606 ATCCAGCAGCTCTGGGAGCCCAAGACAGAGTGGATGACA-AGTCAGGAGATCGAGACCAT 6548
QY 1936 CTTGGGCAACATAGGGGATCCCATCTCTACACACAAAAAATTTTTTAATGAACCCAGC 1995
DB 6547 CCGGCCCAACATAGTGAAGACCTGTCTCTAC-----TAAAAATACAAAAATTAGCCAGGC 6493
QY 1996 ATTGTGGCATGGCTATAGTCCAGCCACTCAAGAGCAGCAGCGGGAGGATCATCTGA 2055
DB 6492 ATATGGGCTGTGGCTGTAGTCCCATCTCAGAGGCTGAGACAGGAGATCGCTTGA 6433
QY 2056 GCCTGGGAGTTGGTTGCAGTGGCTGTATGTACACTGCAGCTCCAGCTGGGCAA 2115
DB 6432 ACACAGGAGCGGAGTTGCAGTGGCCAGATCAGCCACTGCAGCTCCAGCTGGGCAA 6373
QY 2116 CAGACCAAGACCTGTCTCAAAAAATAAACAACATAAATAAATAAATAA 2162
DB 6372 CAGACGAGACCTGTCTCAAAAAATAAACAACATAAATAAATAAATAA 6326

RESULT 13

US-08-257-963B-9/c
Sequence 9, Application US/08257963B
Patent No. 5840686
GENERAL INFORMATION:
APPLICANT: Chader, Gerald J.; Becerra, S.
APPLICANT: Patricia; Schwartz, Joan P.;
APPLICANT: Taniwaki, Takayuki
TITLE OF INVENTION: PIGMENT EPITHELIUM
TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION OF ITS NOVEL
TITLE OF INVENTION: BIOLOGICAL ACTIVITY AND SEQUENCES ENCODING
TITLE OF INVENTION: AND EXPRESSING THE PROTEIN
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morgan & Finnegan
STREET: 345 Park Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,963B
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/952,796
FILING DATE: 24-SEPT-1992
ATTORNEY/AGENT INFORMATION:
NAME: DOROTHY R. AUTH
REGISTRATION NUMBER: 36434
REFERENCE/DOCKET NUMBER: 20264126US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 4421 Base Pairs
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Unknown
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:

OTHER INFORMATION: fragment derived from

Query Match	6.1k	Score	296.6	DB 2	Length	2923
Best Local Similarity	71.3k	Pred. No.	1.4e-45			
Matches	463	Conservative	0	Mismatches	174	Indels
					12	Gaps
QY	1507	AAGATAAGGAGAGCCAGGCGCTGTGGCTCAGCGCTGTAAATCCCGAGCACTTTTAAGAGGCC	1566			
Db	1831	AAAAAGGGGAGCGCTGGGTGTGATAGCTACGCGCTGTAAATCTCAGCACTTTGGGAGCAC	1772			
QY	1567	AAGCGGATGGATCAGCTTGAGCCCGAGGAGTTTCAACACAGCGCTCAGCAACATGTTAAAC	1626			
Db	1771	AAGGAAGGTGGATCAGCTTGAGGTCAGGAGTTTCAAGACCAGCGCTGGCCAAATATGGTGA	1712			
QY	1627	CCCATCTCTACCAAAATACAAAATTAGCTGGCTTGGTGGCTGGCGCTGTATCTCCA	1686			
Db	1712	CCTGCTCTCTACTAAAAATACAAAATTAGCCAGCGGTAGTGGCACACACCTGTAAATCCA	1653			
QY	1687	GCTACTCAGGAGACTGAGCGAAGAGGATCGGTTGAACCTGGGAGCGAGGTTGCAAGTGA	1746			
Db	1652	GCTACTTTGGAGGCTGGGTGGGAGATCACTTGAACCCAGGAGGTCAGGTTGTAGTGA	1593			

This Page Blank (uspto)

Result No.	Score	Query		Length	DB	ID	Description
		Match	Length				
1	1080	54.2	359	1	W69598		Mouse G-protein co
2	908.5	45.6	369	1	W69597		Mouse G-protein co
3	530	26.6	399	1	R66922		Murine C140 recept
4	530	26.6	399	1	W01954		Murine C140 recept
5	525	26.4	395	1	R66920		Murine C140 recept
6	522	26.2	395	1	W01952		Murine C140 recept
7	521.5	26.2	374	1	W51406		Human protease-act
8	508	25.5	398	1	R66921		Human C140 recepto
9	508	25.5	398	1	W01953		Human C140 recepto
10	507.5	25.5	369	1	W51405		Mouse protease-act
11	505.5	25.4	394	1	W51408		Human protease-act
12	494	24.8	397	1	R66923		Human C140 recepto
13	494	24.8	397	1	W01955		Human C140 recepto
14	473	23.8	425	1	R27240		Human thrombin rec
15	473	23.8	425	1	W51407		Human protease-act
16	472	23.7	425	1	R66998		Fragment of the hu
17	472	23.7	892	1	W16314		Human thrombin rec
18	378.5	19.0	322	1	R48715		G-protein coupled
19	378.5	19.0	322	1	W02687		G-protein coupled
20	374	18.8	328	1	R12125		Human placenta G-p
21	354.5	17.8	339	1	R53752		Seven transmembran
22	354.5	17.8	339	1	W07617		Human G-protein th
23	354.5	17.8	339	1	R48733		Human R12 seven tr
24	353.5	17.8	328	1	R91224		Mouse pancreas G-p
25	348	17.5	328	1	W03433		Human placenta pur
26	343	17.2	380	1	R53750		Seven transmembran
27	343	17.2	380	1	W48731		Human R20 seven tr
28	341.5	17.2	370	1	W62597		Human 7-transmembr
29	337.5	17.0	361	1	R54080		Epstein Barr virus
30	337.5	17.0	361	1	W53623		Epstein Barr virus
31	332.5	16.7	355	1	R28272		Sequence in a high
32	332.5	16.7	355	1	R80950		Recombinant high a
33	331.5	16.6	304	1	R48714		G-protein coupled
34	331.5	16.6	304	1	W02686		G-protein coupled
35	330.5	16.6	344	1	W04247		Human G-protein co
36	329.5	16.5	427	1	W61355		Human hypothalamic
37	325	16.3	365	1	W23606		Human p2Y4 recepto
38	323.5	16.2	418	1	R39263		Human somatostatin
39	323.5	16.2	350	1	R68811		Interleukin-8 rece
40	323.5	16.2	1060	1	R70123		IL8-R type 1-GBP 1
41	323.5	16.2	350	1	R80756		Interleukin 8 rece
42	323.5	16.2	350	1	R80951		Recombinant high a
43	322.5	16.2	350	1	R27791		Interleukin-8 rece

```
QY 270 TLAASGRYGHALRLTAVVLASAVAFVPSNLLLLHYSDPSPSANGNLYGAYVPSLALS 329
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 245 ALAANGORYSHALRLTALVLFSAVAFVPSNLLVLYSNPSPANGNLYGAYVPSLALS 304
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 330 TLNSCVDPIIYVVSAREKDKVRAGLQFQSPGDTVASKASAEAGSGRGMTHSSLL 384
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 305 TLNSCVDPIIYVVSAREKDKVR-MLCROPEASSSSQASREAGSRGTAICSSTL 358
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|

RESULT 2
ID W69597 standard; Protein: 360 AA.
AC W69597;
DT 16-OCT-1998 (first entry)
DE Mouse G-protein coupled receptor 69A08 #1 protein.
KW Mouse; CKDLR201.1; chemokine; G-protein coupled receptor; GPCR;
KW 7 transmembrane receptor; inflammation; asthma; antiviral;
KW abnormal cell proliferation; regeneration; degeneration; atrophy.
OS Mus sp.
PN W09831810-A2.
PD 23-JUL-1998.
PF 20-JAN-1998; U00218.
PR 21-JAN-1997; US-786624.
PA (SCHE ) SCHERING CORP.
PI Capone M, Gorman DM, Hedrick JA, Huffine CF, Rossi DL,
PI Vicari A, Zlotnik A;
DR WPI: 98-414108/35.
DR N-PSDB: V40372.
PT Pure or recombinant chemokine CKDLR201.1 - useful, e.g. for treating
PT inflammation and as antiviral agents
PS Claim 2; Page 64-65; 77pp; English.
CC The present sequence is mouse G-protein coupled receptor 69A08 #1
CC protein which comprises a plurality of epitopes. Host cells containing
CC vectors comprising a nucleotide sequence encoding the protein are used to
CC produce recombinant protein. Treatment of a cell, particularly neuron,
CC macrophage or lymphocyte, carrying a G-protein coupled receptor (GPCR)
CC or a receptor responsive to CKDLR201.1 chemokine protein, with an
CC (antagonist is used to control physiological development, e.g.
CC alteration of calcium ion influx, a chemottractant response, morphology,
CC phosphoinositide lipid turnover or an antiviral response. Nucleotide
CC sequences the CKDLR201.1 protein, or GPCRs, are useful as primers or
CC probes, e.g. for detecting and isolating related sequences and for
CC expressing antigenic peptides. Antibodies (Ab) directed against the
CC CKDLR201.1 protein and GPCRs are used to detect or purify the proteins;
CC diagnostically (e.g. for developmental abnormalities); in screening for
CC potential drugs; to inhibit chemokine/receptor activation; (when coupled
CC to a toxin or radioisotope) for killing specific cells, and to raise
CC anti-idiotype antibodies. CKDLR201.1 protein and GPCRs and compounds
CC which bind them can be used to treat inflammation, e.g. asthma; as
CC antiviral agents, and to treat abnormal cell proliferation, regeneration,
CC degeneration and atrophy. Therapeutic agents are administered orally, by
CC injection and rectally.
SQ Sequence 360 AA.

Query Match 45.6%; Score 908.5; DB 1; Length 360;
Best Local Similarity 65.1%; Pred. No. 1.5e-82;
Matches 188; Conservative 17; Mismatches 41; Indels 43; Gaps 3;

QY 117 NLATADLLALALPPRIATHLGRQRPFGACRLATAALYGHMYSVLLLAASVLDRL 176
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 93 NLAVADLLALVLPPLAYHLRGQRPFGACRVATAALYGHMYSVLLLAASVLDRL 152
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 177 ALVHPLARALGRRLALGLCMAAALALPLTLQRTFLARSDRVLC HDALPLDA 236
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 153 ALVHPLARALGRQRTTGLCLVAVLSAATLALPLHLHQNFLL----- 198
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 237 QASHWOPAFCTCALGC-----FPLLAMLLCYGATLHTLAASG 275
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 198 -----APIACCVMMRCPLWSPRTPTGERPSSAWLSAAALPLLAMGLCYGTTLALAANG 251
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 276 RYGHALRLTAVVLASAVAFVPSNLLLLHYSDPSPSANGNLYGAYVPSLALS 335
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
```

```
Db 252 QRYSHALRLTALVLFSAVAFVPSNLLVLYSNPSPANGNLYGAYVPSLALS 311
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 336 DPFIIYYVSAREKDKVRAGLQFQSPGDTVASKASAEAGSGRGMTHSSLL 384
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 312 DPFIIYYVSAREKDKVR-MLCROPEASSSSQASREAGSRGTAICSSTL 359
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|

RESULT 3
ID R66922 standard; Protein: 399 AA.
AC R66922;
DT 22-AUG-1995 (first entry)
DE Murine C140 receptor deduced from cDNA.
KW G-protein-coupled receptor; G-protein; C140 receptor.
OS Mus musculus.
PN W09503318-A.
PD 02-FEB-1995.
PF 26-JUL-1994; U08536.
PR 26-JUL-1993; US-097938.
PA (CORT-) COR THERAPEUTICS.
PI Scarborough RM, Sundelin J;
DR WPI: 95-075182/10.
DR N-PSDB: Q84539.
PT New DNA encoding recombinant C140 receptor - and novel agonists
PT and antagonists and specific antibodies with therapeutic and
PT diagnostic applications.
PS Example: Fig 10; 57pp; English.
CC A cDNA library from a mouse stomach was constructed in lambda gt10
CC and screened with a probe encompassing the C140 genomic clone (see
CC Q84537). A single phage clone was isolated and cut with EcoRI. The
CC insert was cloned in pBluescript and pGS5 and sequenced. The
CC complete nt sequence and deduced AA sequence is given in Q84539 &
CC R66922. 5' RACE resulted in the addition of only 27 bps to the 5'
CC end. The 5' end of the apparent coding region differs from the 5'
CC end of the ORF of genomic DNA; it is believed that the 5' end of
CC the cDNA sequence is correct.
SQ Sequence 399 AA;

Query Match 26.6%; Score 530; DB 1; Length 399;
Best Local Similarity 34.1%; Pred. No. 8.8e-45;
Matches 131; Conservative 68; Mismatches 141; Indels 44; Gaps 11;

QY 13 GFSLSGGTQT-PSVYDESGSTGGDDSTPSILPAPRGYPGQVCANDSDTLELPSSRALL 71
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 37 GBSLIGRLTETPPI-----TGKG-----VPVEPGF-----SIDEFSASIL 71
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 72 LGWPTRLVPALYGLVVLVGPANGALMV-LATQAPRLPSTMLMNLATADLLALALP 130
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 72 TGLKATVLPVVIIVFVIGLPSNGMALWIFLRTKKRPAVYMANLADLLSVIWF 131
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 131 PRIAYHLGRQRPFGACRLATAALYGHMYSVLLLAASVLDRLALVHPL---RARAL 187
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 132 LKISYHLHGNMNVYGEALCKVLIGFFYGNMYCSILFMFCLSVQRYVWVNPMPHKKKA- 191
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 188 RGRRLALGLCMAAALALPLTLQRTFLARSDRVLC HDALPLDAQASHWOPAFCT 247
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 191 ---NTAVGSLALWLIIFLYIPLYVMKQTIYPALNITTC HDVLPVEVLGDMFENYLS 247
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 248 LALLCGCL-PLLAMLLCYGATLHTLAASG-----RRYGHALRLTAVVLASAVAFVPSN 300
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 248 LA-IGVFLFPALLTASAYVLMIKTLRSSAMDEHSEKKRQRAIRLITITLAMYFICFADPN 306
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 301 LLLLHYSDPSPSANGNLYGAYVPSLALS 329
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 307 LLLVHYFLITQRQSHVYALYLVALCULNSCIDPFVIYFVSKDFRDHARNALLCRSV 366
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
QY 360 -PGDTVASKASAEAGSGRGMTHSS 382
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|
Db 367 RTVNRMQISSLSSNKFSSKSGSYSS 390
      |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:| |||:|:|

RESULT 4
```


Db 68 TGLKTTVFLPVVYIIIVEIGLPSNGMALWIFLRTKKKHPAVIYMANLADLLSVIWP 127
 QY 131 PRIATYHLRGORWPFGEAACLTAALYGHMYGVSLLAAVSLDRYLALVHPL---RARAL 187
 Db 128 LKISYHLHGNWYIGALCKYGLGFFYGNMYCSILFMTCLSVQRYWVIWVPMGHPKRKA- 187
 QY 188 RGRRLALGLCMAALMAALALPLTQRTQFRLARSDRVLCHDALPLDAQASHWQPAFTC 247
 Db 187 ---NIAVGSLAIWLLIFLVTIPLYVMKQTIYPALNITTCVDLPEEVLGDMFNFLS 243
 QY 248 LALLGCFPL-LIAMLICYGATLHTLAAG-----RRYGHALRLTAVVLASAVAFVPSN 300
 Db 244 LA-IGVFLFPALLTASAYVLMIKTLRSSAMDEHSENKRQRAIRLIITVLAMYFICFRPSN 302
 QY 301 LLLLLHYDPSPSAWNLYGAYVPSLALSTLNSCDVPFIYIYVSAEFRDKVRAGLFORS- 360
 Db 303 LLLVHYFLIKTQROSHVIALYLYVALCULSTLNSCIDPFVIYFVSKDFRDHARNALLCRSV 362
 QY 360 -PGDTVASKASAEGRSGMGTHSS 382
 Db 363 RTVNRMQISLSSNKFSSKSGSYSS 386

RESULT 6

W01952
 ID W01952 standard; Protein; 395 AA.

AC W01952;

DT 01-APR-1997 (first entry)

DE Murine C140 receptor, including putative signal sequence.

KW C140 receptor; G-protein linked; coupled; seven pass; agonist;

KW antagonist; hypertension; hypotension; blood pressure.

OS Mus sp.

FH Key Location/Qualifiers

FT peptide 1..27

FT /note= "putative signal peptide, differs from

FT signal peptide encoded by a cDNA clone of

FT this receptor (see W01954), the signal

FT sequence given for the cDNA clone is

FT believed to be the correct sequence"

FT protein 28..395

FT /note= "mature protein"

FT modified_site 29

FT /note= "potential Asn-linked glycosylation site"

FT cleavage_site 34..35

FT /note= "putative protease receptor cleavage site"

FT region 78..100

FT /note= "transmembrane region I"

FT region 108..128

FT /note= "transmembrane region II"

FT region 148..169

FT /note= "transmembrane region III"

FT region 188..210

FT /note= "transmembrane region IV"

FT modified_site 220

FT /note= "potential Asn-linked glycosylation site"

FT region 244..264

FT /note= "transmembrane region V"

FT region 286..306

FT /note= "transmembrane region VI"

FT region 324..345

FT /note= "transmembrane region VII"

PN W09623225-Al.

PD 01-AUG-1996.

PF 25-JAN-1996; U01179.

PR 25-JAN-1995; US-390301.

PA (CORT-) COR THERAPEUTICS INC.

PI Scarborough RM, Sundell J.

DR WPI; 96-362813/36.

DR N-PSDB; T32036.

PT Vector for expression C140 cell surface receptor in host cell

PT useful to identify C140 agonist and antagonists, which are

PT antihypertensives and elevators of blood pressure, respectively

PS Example 1; Fig 1A-B; 60pp; English.

CC W01952 represents the murine C140 receptor (C140R), including a putative
 CC signal peptide (see features table). DNA encoding C140R may be
 CC engineered so as to allow the recombinant expression of C140R in a
 CC suitable host cell, i.e. by removing the native expression-control
 CC sequences and replacing them with control sequences operable in the host.
 CC Such a recombinant receptor can be expressed on the surface of oocytes,
 CC this provides a good assay system for identifying agonists/antagonists
 CC of C140R. The C140 receptor is a G-protein linked receptor and a member
 CC of the "seven-pass" transmembrane receptor superfamily (peptide chain
 CC of the receptor passes through the cell membrane seven times, producing
 CC seven transmembrane regions within the receptor molecule). The C140
 CC receptor is involved in controlling blood pressure. C140 antagonists
 CC (see W01942-W01951) are useful to inhibit signalling from this
 CC receptor, resulting in an increase in blood pressure and are therefore
 CC useful in pharmaceuticals for the treatment of hypotension (low blood
 CC pressure). Conversely agonists (see W01914-W01941) of C140 are useful
 CC in pharmaceuticals for the treatment of hypertension (high blood
 CC pressure).
 CC Sequence 395 AA;

Query Match 26.2%; Score 522; DB 1; Length 395;

Best Local Similarity 34.1%; Pred. No. 5.4e-44;

Matches 131; Conservative 66; Mismatches 143; Indels 44; Gaps 11;

QY 13 GFSLSGGTQT-PSVYDESGSTGGDDSTPILPAPRGYPGOVCANDSDTLELPDSSRALL 71

Db 33 GRSILGRLETPPI-----TGKG-----VPVEPGF-----SIDESASIL 67

QY 72 LGWVTRLPALYGLVYVGLPANGALWV-LAQAPRLPSTMLMLNLTADLLALALP 130

Db 68 TGLKTTVFLPVVYIIIVEIGLPSNGMALWIFLRTKKKHPAVIYMANLADLLSVIWP 127

QY 131 PRIATYHLRGORWPFGEAACLTAALYGHMYGVSLLAAVSLDRYLALVHPL---RARAL 187

Db 128 LKISYHLHGNWYIGALCKYGLGFFYGNMYCSILFMTCLSVQRYWVIWVPMGHPKRKA- 187

QY 188 RGRRLALGLCMAALMAALALPLTQRTQFRLARSDRVLCHDALPLDAQASHWQPAFTC 247

Db 187 ---NIAVGSLAIWLLIFLVTIPLYVMKQTIYPALNITTCVDLPEEVLGDMFNFLS 243

QY 248 LALLGCFPL-LIAMLICYGATLHTLAAG-----RRYGHALRLTAVVLASAVAFVPSN 300

Db 244 LA-IGVFLFPALLTASAYVLMIKTLRSSAMDEHSENKRQRAIRLIITVLAMYFICFRPSN 302

QY 301 LLLLLHYDPSPSAWNLYGAYVPSLALSTLNSCDVPFIYIYVSAEFRDKVRAGLFORS- 360

Db 303 LLLVHYFLIKTQROSHVIALYLYVALCULSTLNSCIDPFVIYFVSKDFRDHARNALLCRSV 362

QY 360 -PGDTVASKASAEGRSGMGTHSS 382

Db 363 RTVNRMQISLSSNKFSSKSGSYSS 386

RESULT 7

W51406

ID W51406 standard; Protein; 374 AA.

AC W51406;

DT 12-OCT-1998 (first entry)

DE Human protease-activated receptor 3 (PAR3).

KW Protease-activated receptor 3; PAR3; thrombin receptor; mouse;

KW G-protein coupled receptor; agonist; antagonist; thrombosis;

KW atherosclerosis; restenosis; inflammation; blood coagulation;

KW blood clotting; heart attack; stroke; wound healing;

KW adult respiratory distress syndrome; glomerulosclerosis.

OS Homo sapiens.

FH Key Location/Qualifiers

FT modified_site 25..27

FT /note= "Asn is N-glycosylated"

FT cleavage_site 38..39

FT /note= "thrombin cleavage site"

FT peptide 48..51

FT /note= "hirudin-like sequence"

QY	171	SLDRYLALVHPLRARALRRRLALGLGCMANLWMAALALPLTLQRTFFRLARSDRVLC	230
DB	188	SINRYLAIVHPFTRGLPKHTYALVTCGLVWATVFLYMLPFFFLKQEYLYVQDPITTC	247
QY	231	ALPLDAQASHWQ-PATFCTALLGCEPLPLAMLLCYCATLNTLAASGRRYGHALRLTAVL	289
DB	248	VHNTCESSPFQYFYFISLAFFGLPFFVLIICYAAITLNLNAYDHRWLVYKASLLIL	307
QY	290	ASAVAFVPSGNLLLLHYSDPSAGNLYGAYVPSLALSTLNSCYDPIFYIVYS	344
DB	308	VITFCFAPSNIILIIHANYNNNDGLYFVLYIALCLGSLNSCLDPFLYFLMS	362

RESULT

8

R66921

R66921 standard; Protein; 398 AA.

ID

AC

R66921;

DT

22-AUG-1995 (first entry)

DE

Human C140 receptor.

KW

G-protein-coupled receptor; G-protein; C140 receptor.

OS

Homo sapiens.

FH

Key

FT

Location/Qualifiers

FT

protein

FT

1. .27

FT

/label= signal peptide

FT

31

FT

/label= Asn linked glycosylation site

FT

223

FT

/label= Asn linked glycosylation site

FT

37. .38

FT

/label= protease receptor cleavage site

FT

81.103

FT

/label= transmembrane I

FT

11.132

FT

/label= transmembrane II

FT

150.174

FT

/label= transmembrane III

FT

191.212

FT

/label= transmembrane IV

FT

245.267

FT

/label= transmembrane V

FT

289.309

FT

/label= transmembrane VI

FT

327.348

FT

/label= transmembrane VII

PN

W09503318-A.

PD

02-FEB-1995.

PF

26-JUL-1994; U08536.

PR

26-JUL-1993; US-097938.

PA

(COR-) COR THERAPEUTICS.

PI

Scarborough RM, Sundelln J;

DR

WFI; 95-075182/10.

DR

N-PSDB; Q84558.

PT

New DNA encoding recombinant C140 receptor - and novel agonists

PT

and antagonists and specific antibodies with therapeutic and

PT

diagnostic applications.

PS

Disclosure; Fig 2; 57pp; English.

CC

The availability of genomic DNA encoding the mouse protease

CC

C140 receptor (see Q84557) permitted the retrieval of the corresp.

CC

human gene. A human genomic library cloned in the vector EMBL3 was

CC

screened using the entire coding region of the murine clone as a

CC

probe. The recovered human gene including the DNA sequence and the

CC

deduced AA sequence are shown in Q84558 & R66921. Subsequent

CC

experiments indicated that the human C140 gene is located in the

CC

same region of the long arm of chromosome number 5 (5q12-5q13)

CC

as has been reported for the human thrombin receptor gene.

SQ

Sequence 398 AA;

Query Match

25.5%; Score 508; DB 1; Length 398;

Best Local Similarity

36.5%; Pred. No. 1.4e-42;

Matches 110; Conservative

57; Mismatches 124; Indels 10; Gaps

Best Local Similarity	50.58	Freq. NO.	1.48	42	
Matches	110	Conservative	57	Mismatches	124
				Indels	10
				Gaps	5

QY 67 SRALLGWPTRLVPALYGLVLVGLPANGALWV-LATQAPRLPSTMLLMNLATADLLL 125

Db 66 SASVLTGKLTTFVLPVYIVFVVGSLPNSGMALWFLRTKKKHPAVIYMANLADLLS 125
 QY 126 ALALPPRIAYHLRGQWPFGAEACRLATAALYGHMYGVSLLLAANVSLDRYLALVHPLRAR 185
 Db 126 VWFPKIAIYHNGNWIYGEALCNVLIGFFYGNMYCSILFTCLSVQRYWVWVNP-CH 184
 QY 186 ALRGRLALGLCMAWMLAALPLTLQRTQFRRLARSRVLCCHDALPDAQASHWQPAF 245
 Db 185 SRKKANIAIGISLAIWLLILLITPIYVVKQTFIPALNITTCCHDVLPPQLLVGMFNYF 244
 QY 246 TCLALLGCFL-PLLAMLLCYGATLHTLAASG-----RRYGHALRLTAVVLASAVAFVFP 298
 Db 245 LSLA-IGVFLPAPFLTASAYVLMIRLSAMDENSEKKRKAIRKLIIVTVLAMYLICFTP 303
 QY 299 SNLLLLHYSDSPSANGNLYGAYVPSLALSTLNSCVDPIFYIYVSAEFDKVRAGLQFR 358
 Db 304 SNLLLVHYFLIKSGQSHVYALIVLALCLSTLNSCIDPFYIYFVSHDFRDKHAKNALLCR 363
 QY 359 S 359
 Db 364 S 364

RESULT 9

W01953
 ID W01953 standard; Protein: 398 AA.
 AC W01953;
 DT 01-APR-1997 (first entry)
 DE Human C140 receptor, with putative signal sequence.
 KW C140 receptor; G-protein linked; coupled; seven pass; agonist;
 KW antagonist; hypertension; hypotension; blood pressure.
 OS Homo sapiens.
 FH Key
 FT peptide
 FT 1. 27
 FT /note= "putative signal peptide, differs from
 FT signal peptide encoded by a cDNA clone of
 FT this receptor (see W01953), the signal
 FT sequence given for the cDNA clone is
 FT believed to be the correct sequence"
 FT 28. 398
 FT /note= "mature protein"
 FT 31
 FT modified_site
 FT /note= "potential Asn-linked glycosylation site"
 FT cleavage_site
 FT 37. 38
 FT /note= "putative protease receptor cleavage site"
 FT region
 FT 81. 103
 FT /note= "transmembrane region I"
 FT region
 FT 111. 132
 FT /note= "transmembrane region II"
 FT region
 FT 151. 174
 FT /note= "transmembrane region III"
 FT region
 FT 191. 212
 FT /note= "transmembrane region IV"
 FT modified_site
 FT 223
 FT /note= "potential Asn-linked glycosylation site"
 FT region
 FT 245. 267
 FT /note= "transmembrane region V"
 FT region
 FT 289. 309
 FT /note= "transmembrane region VI"
 FT region
 FT 327. 348
 FT /note= "transmembrane region VII"
 FT W09623225-A1.
 PN 01-AUG-1996.
 PD 01-AUG-1996.
 PR 25-JAN-1996; U01179.
 PR 25-JAN-1995; US-390301.
 PA (CORT-) COR THERAPEUTICS INC.
 PI Scarborough RW, Sundelin J;
 DR WPI: 96-362813/36.
 DR N-PSDB: T32037.
 DR Vector for expression C140 cell surface receptor in host cell -
 PT useful to identify C140 agonist and antagonists, which are
 PT antihypertensives and elevators of blood pressure, respectively

PS Example 2; Fig 2A-B; 60pp; English.
 CC W01953 represents the human C140 receptor (C140R), including a putative
 CC signal peptide (see features table). DNA encoding of C140R may be
 CC engineered so as to allow the recombinant expression of C140R in a
 CC suitable host cell, i.e. by removing the native expression-control
 CC sequences and replacing them with control sequences operable in the host.
 CC Such a recombinant receptor can be expressed on the surface of oocytes,
 CC this provides a good assay system for identifying agonists/antagonists
 CC of C140R. The C140 receptor is a G-protein linked receptor and a member
 CC of the "seven-pass" transmembrane receptor superfamily (peptide chain
 CC of the receptor passes through the cell membrane seven times, producing
 CC seven transmembrane regions within the receptor molecule). The C140
 CC receptor is involved in controlling blood pressure. C140 antagonists
 CC (see W01942-W01951) are useful to inhibit signalling from this
 CC receptor, resulting in an increase in blood pressure and are therefore
 CC useful in pharmaceuticals for the treatment of hypotension (low blood
 CC pressure). Conversely agonists (see W01914-W01941) of C140 are useful
 CC in pharmaceuticals for the treatment of hypertension (high blood
 CC pressure).
 CC Sequence 398 AA;
 SQ

Query Match 25.5%; Score 508; DB 1; Length 398;

Best Local Similarity 36.5%; Pred. No. 1.4e-42;

Matches 110; Conservative 57; Mismatches 124; Indels 10; Gaps 5;

QY 67 SRALLGHWPTRLYPALYGLVVLVGLPANGALAV-LATQAPRLPSTMLMLNATADLL 125
 Db 66 SASVLTGKLTTFVLPVYIVFVVGSLPNSGMALWFLRTKKKHPAVIYMANLADLLS 125
 QY 126 ALALPPRIAYHLRGQWPFGAEACRLATAALYGHMYGVSLLLAANVSLDRYLALVHPLRAR 185
 Db 126 VWFPKIAIYHNGNWIYGEALCNVLIGFFYGNMYCSILFTCLSVQRYWVWVNP-CH 184
 QY 186 ALRGRLALGLCMAWMLAALPLTLQRTQFRRLARSRVLCCHDALPDAQASHWQPAF 245
 Db 185 SRKKANIAIGISLAIWLLILLITPIYVVKQTFIPALNITTCCHDVLPPQLLVGMFNYF 244
 QY 246 TCLALLGCFL-PLLAMLLCYGATLHTLAASG-----RRYGHALRLTAVVLASAVAFVFP 298
 Db 245 LSLA-IGVFLPAPFLTASAYVLMIRLSAMDENSEKKRKAIRKLIIVTVLAMYLICFTP 303
 QY 299 SNLLLLHYSDSPSANGNLYGAYVPSLALSTLNSCVDPIFYIYVSAEFDKVRAGLQFR 358
 Db 304 SNLLLVHYFLIKSGQSHVYALIVLALCLSTLNSCIDPFYIYFVSHDFRDKHAKNALLCR 363
 QY 359 S 359
 Db 364 S 364

RESULT 10

W51405

ID W51405 standard; Protein: 369 AA.

AC W51405;

DT 12-OCT-1998 (first entry)

DE Mouse protease-activated receptor 3 (PAR3).

KW Protease-activated receptor 3; PAR3; thrombin receptor; mouse;

KW G-protein coupled receptor; agonist; antagonist; thrombosis;

KW atherosclerosis; restenosis; inflammation; blood coagulation;

KW blood clotting; heart attack; stroke; wound healing;

KW adult respiratory distress syndrome; glomerulosclerosis.

OS Mus sp.

PN W09818456-A1.

PD 07-MAY-1998.

PF 29-OCT-1997; U19732.

PR 30-OCT-1996; US-742440.

PA (REGC) UNIV CALIFORNIA.

PI Connolly A, Coughlin SR, Ishihara H;

DR WPI: 98-271905/24.

DR N-PSDB: V07372.

PT DNA encoding protease-activated receptor 3 - for detection of
 specific agonists and antagonists, potentially useful for treating

PT e.g. thrombosis, atherosclerosis, inflammation etc.
PS Claim 3; Page 39-40; 74pp; English.
CC This polypeptide comprises mouse protease-activated receptor 3
CC (PAR3), a cell surface protein which is specifically activated by
CC thrombin or a thrombin agonist, thereby activating signalling
CC events such as phosphoinositide hydrolysis, calcium ion efflux and
CC platelet aggregation. Its amino acid sequence was deduced from an
CC isolated cDNA clone (see V07372). The human PAR3 amino acid
CC sequence (see W51406) is also provided. Also claimed are vectors,
CC host cells and an assay device. Host cells are used to screen
CC compounds for their ability to act as agonists or antagonists of
CC the effects of thrombin-PAR3 interaction. Agonists are used to
CC treat wounds, thrombosis, atherosclerosis, restenosis, inflammation
CC and other thrombin activated disorders. Antagonists (see W51415-21)
CC are used to control blood coagulation and thereby to treat heart
CC attack and stroke. They also mediate inflammatory and proliferative
CC responses to injury as occur in wound healing, atherosclerosis,
CC restenosis, pulmonary inflammation (ARDS) and glomerulosclerosis.
SQ Sequence 369 AA;

Query Match 25.5%; Score 507.5; DB 1; Length 369;
Best Local Similarity 31.4%; Pred. No. 1.4e-42;
Matches 111; Conservative 76; Mismatches 151; Indels 15; Gaps 6;
QY 5 LLLPLVL---GFLSGGTQTPSYVDESGTGGDDSTPSILPAP--RGYPGV----- 54
DB 11 LLFLVTVCSGGINVSNKAPITIK--SFNGGPQNTFEFLSDIEGWTGATTIKAE 68
QY 54 CANDS-DTLELPDSSRALLGWPTRLVPALYGLVVLVGLPANGALWLAATAPRLPST 112
DB 69 CPESISTLHVNNATIGYLRSSLSQTQVPAIYILVVGVPNSIVTLWKLRLTKSISLV 128
QY 113 MLMLNLTADLLALALPPIAYHLRGQRPFGAARLATAALYGHMYSVLLAAVSL 172
DB 129 IFHTNLATADLEFCVTLPEKIAHNGNWNVGEVCMCRITTVFVGYNNYCAILLTCGI 188
QY 173 DRYALVHLPLARALGRRLALGCMALWMAALPLTLQRTFLARSDRVLCCHDAL 232
DB 189 NRYLATAHPFTYQKLPRKFSLSGIVMVFVLYMLPFVILKQEHVHVHSEITTCDDV 248
QY 233 -PLDAQASHMOPAFCTALLCFLPLAMLCYCATLHTLAASGRYRGHARLTAVALAS 291
DB 249 DACESPSSFRYYVSLAFFGLFPFVFIICVYTLHLKSKDRWLGKAVILLIYI 308
QY 292 AVAFVPSNLLLHYSDPSANGNLYGAYVPSLALSTLNSCVDPFIYYVSAFVKRAGLFORS 344
DB 309 FTICFAPNTIILVIHANYHYHNTDSLYFMYLIALCLGSLNSCLDPLFYFVMS 361

RESULT 11
W51408
ID W51408 standard; Protein; 394 AA.
AC W51408;
DT 12-OCT-1998 (first entry)
DE Human protease-activated receptor 2 (PAR2).
KW Protease-activated receptor 2; PAR2; PAR3; thrombin receptor;
KW human.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Cleavage_site 36..37 /note= "thrombin cleavage site"
FT
PN W09818456-A1.
PD 07-MAY-1998.
PF 29-OCT-1997; U19732.
PR 30-OCT-1996; US-742440.
PA (REGC) UNIV CALIFORNIA.
PI Connolly A, Coughlin SR, Ishihara H;
DR WPI; 98-271905/24.
PT DNA encoding protease-activated receptor 3 - for detection of
PT specific agonists and antagonists, potentially useful for treating
PT e.g. thrombosis, atherosclerosis, inflammation etc.
PS Example 1; Page 43-44; 74pp; English.

CC This polypeptide comprises human protease-activated receptor 2
CC (PAR2). The physiological activator of PAR2 remains unknown;
CC it is not activated by thrombin. The invention relates to novel
CC mouse and human PAR3 (see W51405-06) that show homology to PAR2 and
CC which are specific receptors for thrombin. They can be used to
CC screen for specific agonists and antagonists of thrombin useful
CC e.g. for treating atherosclerosis, thrombosis and inflammation.
SQ Sequence 394 AA;

Query Match 25.4%; Score 505.5; DB 1; Length 394;
Best Local Similarity 34.1%; Pred. No. 2.4e-42;
Matches 122; Conservative 64; Mismatches 151; Indels 21; Gaps 9;
QY 10 LVLGFLSGGTQTPSYVDESGTGGDDSTPSILPAPRGYPGVGCANDSDTLELPDSSRA 69
DB 16 LAASLSCSGTQGTNRSSKSLGKVDGTSHV--TGKGVTVET-----VFSVDEFSAS 67
QY 70 LLLGWPTRLVPALYGLVVLVGLPANGALWV-LATQAPRLPSTMLMLNLTADLLALA 128
DB 68 VLTKGLTVFELPIYITIVFVGLPSNGMALWVLFRTKKKHPAVIYMANLALADLLSVI 127
QY 129 LPPIAYHLRGQRPFGAARLATAALYGHMYSVLLAAVSLDRYLALVHLPLARALR 188
DB 128 FPLKIAHYHNGNWNVGEALCNVLIGFFYGNMYCSILFMTCLSVQRVWVIVNPM-GHSRK 186
QY 189 GRRALGLCMAWMAALPLTLQRTFLARSDRVLCCHDALPDAQASHMOPAFCTCL 248
DB 187 KANIAIGISLAIWLLILLVTPLYVVKQTFIPALNTTCHDVLV--EQLLVGDP-FLSL 243
QY 249 ALLGCFL-PLLAMLCVGLTHTLAASG-----RRYGHALRLTAVVLAASAVFFVPSNL 301
DB 244 A-IGVFLPFAFLASAVLMIRLSRSDENSEKKRRAIKLIVTVLAWLYLCFTFPSNL 302
QY 302 LLLLHYSDPSANGNLYGAYVPSLALSTLNSCVDPFIYYVSAFVKRAGLFORS 359
DB 303 LLVVHFLIISQGSQSHVYALYVALCLSTLNSCIDFVYFVSHDFDHAKNALLCRS 360

RESULT 12
R66923
ID R66923 standard; Protein; 397 AA.
AC R66923;
DT 22-AUG-1995 (first entry)
DE Human C140 receptor encoded by cDNA.
KW G-protein-coupled receptor; G-protein; C140 receptor.
OS Homo sapiens.
PN W09503318-A.
PD 02-FEB-1995.
PF 26-JUL-1994; U08536.
PR 28-JUL-1993; US-097938.
PA (CORT-) COR THERAPEUTICS.
PI Scarborough RM, Sundelin J;
DR WPI; 95-075182/10.
DR N-PSDB; Q84560.
PT New DNA encoding recombinant C140 receptor - and novel agonists
PT and antagonists and specific antibodies with therapeutic and
PT diagnostic applications.
PS Example; Fig 11; 57pp; English.
CC A human intestinal tumour cDNA library was subjected to PCR using
CC primers designed from the genomic clone (see Q84558) and the
CC amplified fragment was cloned in pSG5 and sequenced. There are
CC four AA differences between the cDNA encoded sequence and that
CC encoded by the genomic DNA. The genomic DNA sequence and deduced
CC AA sequence are given in Q84560 & R66923.
SQ Sequence 397 AA;

Query Match 24.8%; Score 494; DB 1; Length 397;
Best Local Similarity 32.7%; Pred. No. 3.4e-41;
Matches 117; Conservative 64; Mismatches 159; Indels 18; Gaps 7;
QY 10 LVLGFLSGGTQTPSYVDESGTGGDDSTPSILPAPRGYPGVGCANDSDTLELPDSSRA 69

```
Db 16 LAASLSCSGTIOGTRSSKGRSLIGKVDGTSVH--TGKGVTVET-----VFSVDEFSAS 67
QY 70 LLLGWTPRLVPALYGLVVLVGPANGALWV-LATQAPRLPSTMLMNLATADLLALA 128
Db 68 VLAKGTTTFLPITVYIVFAVGLPSNGMALWFLFKKKHPPAVIYMANLALADLLSVI 127
QY 129 LPPIAYHLRGQWPFGAEACRLATAALYGHMYGVSLLAAVSLDRYLAALVHPLRARALR 188
Db 128 FPKIAYHIGNNWIYGEALCNVLIGFFYNNMYCSILFTCLSVQRYWIVNPM-GHSRK 186
QY 189 GRRALGLCMAAWMAALALPLTLQRTFLARSDRVLCHEALDPAQASHWQPAFTCL 248
Db 187 KANIAIGISLAIWLLTLLVTPLYVVKQTFIPALNITTCCHDVLPEQLLVGDMFNFLSL 246
QY 249 ALLGCFL-PLLAMLLCYGATLHTLAAG-----RRYGHALRLTAVVLASAVAFVPSNL 301
Db 247 A-IGVELFPAFLTASAYVLMIRMLRSSAMDENSEKRRKRAIKLIVTVLGMYLICFTPSNL 305
QY 302 LLLHYSDPSPSANGNLYGAYVPSLALSTLNSCVDPFIYVYVSAEFRDKVRAGLFORS 359
Db 306 LLVWHFLIKSQGSHVYALYVALCLSTLNSCIDPFYIFVSHDRNAKNALLCRS 363

RESULT 13
W01955
ID W01955 standard; Protein; 397 AA.
AC W01955;
DT 02-APR-1997 (first entry)
DE Human C140 receptor.
KW C140 receptor; G-protein linked; coupled; seven pass; agonist;
KW antagonist; hypertension; hypotension; blood pressure.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide
FT 1..27 /note= "the signal peptide differs from that
FT encoded by a genomic DNA sequence for
FT this receptor (see W01953), the signal
FT sequence given here is believed to be
FT the correct sequence"
FT protein
FT 28..397
FT /note= "mature protein"
PN W09623225-A1.
PD 01-AUG-1996.
PF 25-JAN-1996; U01179.
PR 25-JAN-1995; US-390301.
PA (CORT-) COR THERAPEUTICS INC.
PI Scarborough RM, Sundell J.
DR WPI: 96-362813/36.
DR N-PSDB; T32039.
PT Vector for expression C140 cell surface receptor in host cell -
PT useful to identify C140 agonist and antagonists, which are
PT antihypertensives and elevators of blood pressure, respectively
PT Example 5; Fig 11A-B; 60pp; English.
PS W01955 represents the human C140 receptor (C140R). DNA encoding C140R
CC may be engineered so as to allow the recombinant expression of C140R in
CC a suitable host cell, i.e. by removing the native expression-control
CC sequences and replacing them with control sequences operable in the host.
CC Such a recombinant receptor can be expressed on the surface of oocytes,
CC this provides a good assay system for identifying agonists/antagonists
CC of C140R. The C140 receptor is a G-protein linked receptor and a member
CC of the "seven-pass" transmembrane receptor superfamily (peptide chain
CC of the receptor passes through the cell membrane seven times, producing
CC seven transmembrane regions within the receptor molecule). The C140
CC receptor is involved in controlling blood pressure. C140 antagonists
CC (see W01942-W01951) are useful to inhibit signalling from this
CC receptor, resulting in an increase in blood pressure and are therefore
CC useful in pharmaceuticals for the treatment of hypotension (low blood
CC pressure). Conversely agonists (see W01914-W01941) of C140 are useful
CC in pharmaceuticals for the treatment of hypertension (high blood
CC pressure).
SQ Sequence. 397 AA;
```

```
Query Match 24.8%; Score 494; DB 1: Length 397;
Best Local Similarity 32.7%; Pred. No. 3.4e-41;
Matches 117; Conservative 64; Mismatches 159; Indels 18; Gaps 7;

QY 10 LVLGFLSGGTQTPSVYDESGTGGDDSTPSILPAPRGYPGQVCANDSTLELPDSSRA 69
Db 16 LAASLSCSGTIOGTRSSKGRSLIGKVDGTSVH--TGKGVTVET-----VFSVDEFSAS 67
QY 70 LLLGWTPRLVPALYGLVVLVGPANGALWV-LATQAPRLPSTMLMNLATADLLALA 128
Db 68 VLAKGTTTFLPITVYIVFAVGLPSNGMALWFLFKKKHPPAVIYMANLALADLLSVI 127
QY 129 LPPIAYHLRGQWPFGAEACRLATAALYGHMYGVSLLAAVSLDRYLAALVHPLRARALR 188
Db 128 FPKIAYHIGNNWIYGEALCNVLIGFFYNNMYCSILFTCLSVQRYWIVNPM-GHSRK 186
QY 189 GRRALGLCMAAWMAALALPLTLQRTFLARSDRVLCHEALDPAQASHWQPAFTCL 248
Db 187 KANIAIGISLAIWLLTLLVTPLYVVKQTFIPALNITTCCHDVLPEQLLVGDMFNFLSL 246
QY 249 ALLGCFL-PLLAMLLCYGATLHTLAAG-----RRYGHALRLTAVVLASAVAFVPSNL 301
Db 247 A-IGVELFPAFLTASAYVLMIRMLRSSAMDENSEKRRKRAIKLIVTVLGMYLICFTPSNL 305
QY 302 LLLHYSDPSPSANGNLYGAYVPSLALSTLNSCVDPFIYVYVSAEFRDKVRAGLFORS 359
Db 306 LLVWHFLIKSQGSHVYALYVALCLSTLNSCIDPFYIFVSHDRNAKNALLCRS 363

RESULT 14
R27240
ID R27240 standard; Protein; 425 AA.
AC R27240;
DT 25-MAY-1994 (second entry)
DT 15-FEB-1993 (first entry)
DE Human thrombin receptor (TR).
KW Diagnosis; cardiovascular disease; wound healing; restenosis;
KW thrombosis; unstable angina treatment; myocardial infarction;
KW thrombotic; thromboembolytic stroke.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide
FT 1..26 /label= sig_peptide
FT 28..425 /label= mat_protein
FT 28..41
FT /note= "activation peptide RPESKATNATLDPR"
FT cleavage_site 41..42
FT /note= "thrombin-catalysed cleavage-site"
FT disulfide_bond 175..254
FT /note= "analogous to rhodopsin and beta-2
FT adrenergic receptor"
FT domain 100..129
FT /note= "transmembrane domain I"
FT domain 137..160
FT /note= "transmembrane domain II"
FT domain 177..197
FT /note= "transmembrane domain III"
FT domain 219..239
FT /note= "transmembrane domain IV"
FT domain 266..294
FT /note= "transmembrane domain V"
FT domain 311..335
FT /note= "transmembrane domain VI"
FT domain 348..367
FT /note= "transmembrane domain VII"
FT region 34..36
FT /note= "consensus N-linked glycosylation site"
FT region 62..64
FT /note= "consensus N-linked glycosylation site"
FT region 75..77
FT /note= "consensus N-linked glycosylation site"
```

```
FT region 250..252
FT /note= "consensus N-linked glycosylation site"
FT region 259..261
FT /note= "consensus N-linked glycosylation site"
PD WO214750-A.
PN 03-SEP-1992.
PR 19-FEB-1992; U01312.
PR 19-FEB-1991; US-657769.
PR 07-NOV-1991; US-789184.
PA (CORT-) COR THERAPEUTICS INC.
PA (REGC ) UNIV CALIFORNIA.
PI Coughlin SR, Scarborough RM;
DR WPI; 92-316119/38.
DR N-PSDB; Q28568.
PT DNA encoding cell surface receptor for thrombin - useful for
PT determining thrombin in diagnosing e.g. cardiovascular diseases,
PT also to treat wound healing, restenosis etc.
PS Disclosure; Fig 1; 81pp; English.
CC The TR DNA sequence can be used in the prepn. of diagnostics to
CC determine thrombin levels in samples, and screening tools for
CC candidate substances which affect thrombin activity in vivo.
CC Thrombosis may be diagnosed in a mammal by measuring the presence,
CC absence or amt. of the cleaved activation peptide of the TR.
SQ Sequence 425 AA;
```

Query Match 23.8%; Score 473; DB 1; Length 425;
Best Local Similarity 32.1%; Pred. No. 4.6e-39;
Matches 133; Conservative 76; Mismatches 169; Indels 36; Gaps 10;

```
QY 4 RLLWPLVLGFLSG---GTQTPSVYDESGTGGDDSTPSILPAP-----RG 48
Db 5 RLL--VAACFSLCGPLLSARTARRPESKATNATLDPRSFLLRNPNKYEPFWEDEKN 62
QY 49 YPG-----QVCANDSTL--ELP-----DSSRALLLGWPTRLVPALYGLVVLGPANG 96
Db 63 ESLTEYRLVLSINKSPLOKQPAFTSEDASGYLTSSWL-TLFVPSVYTCVFFVSLPLNI 121
QY 97 LAHWV-LATQAPRLPTMLMNLNATADLLALPPRIAYHLRGQRPFGAACLATAA 155
Db 122 MAIVFILKMKVKKPAVVYMLHATADLVFVLPFKISYFSGSDWQFGSELCREVTAA 181
QY 156 LYGHMYGVSLLAAVSLDRYLALVHPLRARALGRRLALGLCMAAALALPLTLQR 215
Db 182 FYCNMYASILLMTVISIDRFALVVPMSQSLWRTLGRASFTCLAIWALAGVVPVLKE 241
QY 216 QTFRLARSDRVLCHDALPLDAQASHWQPAFTCLALGCFPLMLLCLCYGATLHTLAASG 275
Db 242 QTIVPGLNITTCCHDVNLLEGGYAYYFSAFSAVFFVPLIISTVCYVSIIRCLSSA 301
QY 276 ----RRYGHALRLTAVVLASAVAFFVPSNLLLLLHYSDPS-PSAMGNLYGAYVPSLALST 330
Db 302 VANRSKSRALFSLAAVFCIFICFGPTNVLIAHYSLFSLTSTTEAAYFAVLLCVCSV 361
QY 331 LNSCVDPFYIYVVAEFRDKVRAGLQRPFGDTVASKASAEGRSGMGTHSSLL 384
Db 362 ISSCIDPLIYYASSECORYVYSILCKESSDPSSYNSSGQLMASKMDTCSSNL 415
```

RESULT 15
W51407
ID W51407 standard; Protein; 425 AA.
AC W51407;
DT 12-OCT-1998 (first entry)
DE Human protease-activated receptor 1 (PAR1).
KW Protease-activated receptor 1; PAR1; PAR3; thrombin receptor;
KW human.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Binding_site 52..55
FT /note= "thrombin binding site"
FT Cleavage_site 41..42
FT /note= "thrombin cleavage site"

```
PN WO9818456-A1.
PD 07-MAY-1998.
PF 29-OCT-1997; U19732.
PR 30-OCT-1996; US-742440.
PA (REGC ) UNIV CALIFORNIA.
PI Connolly A, Coughlin SR, Ishihara H;
DR WPI; 98-271905/24.
PT DNA encoding protease-activated receptor 3 - for detection of
PT specific agonists and antagonists, potentially useful for treating
PT e.g. thrombosis, atherosclerosis, inflammation etc.
PS Example 1; Page 42-43; 74pp; English.
CC This polypeptide comprises human protease-activated receptor 1
CC (PAR1), a receptor that mediates thrombin signalling. The
CC invention relates to novel mouse and human PAR3 (see W51405-06)
CC that show homology to PAR1 and which are specific receptors for
CC thrombin. They can be used to screen for specific agonists and
CC antagonists of thrombin useful e.g. for treating atherosclerosis,
CC thrombosis and inflammation.
SQ Sequence 425 AA;
```

Query Match 23.8%; Score 473; DB 1; Length 425;
Best Local Similarity 32.1%; Pred. No. 4.6e-39;
Matches 133; Conservative 76; Mismatches 169; Indels 36; Gaps 10;

```
QY 4 RLLWPLVLGFLSG---GTQTPSVYDESGTGGDDSTPSILPAP-----RG 48
Db 5 RLL--VAACFSLCGPLLSARTARRPESKATNATLDPRSFLLRNPNKYEPFWEDEKN 62
QY 49 YPG-----QVCANDSTL--ELP-----DSSRALLLGWPTRLVPALYGLVVLGPANG 96
Db 63 ESLTEYRLVLSINKSPLOKQPAFTSEDASGYLTSSWL-TLFVPSVYTCVFFVSLPLNI 121
QY 97 LAHWV-LATQAPRLPTMLMNLNATADLLALPPRIAYHLRGQRPFGAACLATAA 155
Db 122 MAIVFILKMKVKKPAVVYMLHATADLVFVLPFKISYFSGSDWQFGSELCREVTAA 181
QY 156 LYGHMYGVSLLAAVSLDRYLALVHPLRARALGRRLALGLCMAAALALPLTLQR 215
Db 182 FYCNMYASILLMTVISIDRFALVVPMSQSLWRTLGRASFTCLAIWALAGVVPVLKE 241
QY 216 QTFRLARSDRVLCHDALPLDAQASHWQPAFTCLALGCFPLMLLCLCYGATLHTLAASG 275
Db 242 QTIVPGLNITTCCHDVNLLEGGYAYYFSAFSAVFFVPLIISTVCYVSIIRCLSSA 301
QY 276 ----RRYGHALRLTAVVLASAVAFFVPSNLLLLLHYSDPS-PSAMGNLYGAYVPSLALST 330
Db 302 VANRSKSRALFSLAAVFCIFICFGPTNVLIAHYSLFSLTSTTEAAYFAVLLCVCSV 361
QY 331 LNSCVDPFYIYVVAEFRDKVRAGLQRPFGDTVASKASAEGRSGMGTHSSLL 384
Db 362 ISSCIDPLIYYASSECORYVYSILCKESSDPSSYNSSGQLMASKMDTCSSNL 415
```

Search completed: October 28, 1999, 15:58:42
Job time: 20228 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 1999, 05:16:12 ; Search time 17.19 seconds
(without alignments)
1378.377 Million cell updates/sec

Title: US-09-371-333-2
Perfect score: 1991
Sequence: 1 MWGRLLPLVLGFSLSGGT.....SKASABGGSGMGTHSSLLQ 385

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database : SPTREML10.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1991	100.0	385	4	076067
2	1466	73.6	396	11	088634
3	357	17.9	328	13	093361
4	353	17.7	374	13	057466
5	345.5	17.4	370	4	015132
6	344	17.3	380	6	097866
7	340.5	17.1	345	11	070129
8	336	16.9	371	11	092280
9	332	16.7	361	11	035811
10	321.5	16.1	371	11	088854
11	320	16.1	387	4	043803
12	317.5	15.9	370	11	088626
13	314	15.8	342	13	093239
14	313	15.7	368	4	060755
15	309.5	15.5	337	4	075819
16	309.5	15.5	370	11	054914
17	308.5	15.5	353	13	079960
18	306.5	15.4	370	11	088853
19	304	15.3	371	4	043190
20	300.5	15.1	367	11	088410
21	300	15.1	347	11	088535
22	298.5	15.0	355	6	097664
23	298.5	15.0	383	12	089609
24	298	15.0	352	6	018770
25	297.5	14.9	356	6	097571
26	297.5	14.9	475	11	088680
27	297	14.9	352	6	077776
28	297	14.9	352	6	097975
29	294.5	14.8	358	11	092213

ALIGNMENTS

RESULT 1	30	293.5	14.7	352	6	097883
076067	31	293	14.7	352	4	015538
ID	076067	PRELIMINARY;				PRT; 385 AA.
AC	076067					
DT	01-NOV-1998	(TREMBLrel. 08, Created)				
DT	01-NOV-1998	(TREMBLrel. 08, Last sequence update)				
DT	01-NOV-1998	(TREMBLrel. 01, Last annotation update)				
DE	PROTEASE-ACTIVATED RECEPTOR 4.					
OS	Homo sapiens (Human)					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;					
OC	Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RA	KAHN M.L., ZHENG Y.W., HUANG W., BIGORNIA V., ZENG D., MOFF S.,					
RT	"Gene and locus structure and chromosomal localization of the					
RT	protease-activated receptor gene family.";					
RL	J. Biol. Chem. 0:0-0(1998).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RA	KAHN M.L., ZHENG Y.W., HUANG W., BIGORNIA V., ZENG D., MOFF S.,					
RT	"A dual thrombin-receptor system for platelet activation.";					
RL	Nature 0:0-0(1998).					
RN	[3]					
RP	SEQUENCE FROM N.A.					
RA	XU W., ANDERSEN H., WHITMORE T.E., PRESNELL S.R., YEE D.P., CHING A.,					
RT	GILBERT T., DAVIE E.W., FOSTER D.C.;					
RT	"Cloning and characterization of human protease-activated receptor					
RT	4.";					
RL	Proc. Natl. Acad. Sci. U.S.A. 95:6642-6646(1998).					
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).					
CC	-1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.					
DR	EMBL; AF080214; AAC28860.1; -					
DR	EMBL; AF055917; AAC25699.1; -					
DR	PFAM; PF00001; 7tm1.1; -					
DR	PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.					
KW	Protease; G-protein coupled receptor; Transmembrane; Glycoprotein.					
SQ	SEQUENCE 385 AA; 41162 MW; 21A9FC96 CRC32;					

Query Match	100.0%;	Score	1991;	DB	4;	Length	385;
Best Local Similarity	100.0%;	Pred. No.	1.8e-142;				
Matches	385;	Conservative	0;	Mismatches	0;	Indels	0;
Matches	385;	Conservative	0;	Mismatches	0;	Indels	0;
QY	1	MWGRLLPLVLGFSLSGGTQTFSVYDES	CGTSTGGDDSTFSLPAPG	YGCVCANDSDT	60		
DB	1	MWGRLLPLVLGFSLSGGTQTFSVYDES	CGTSTGGDDSTFSLPAPG	YGCVCANDSDT	60		
QY	61	LELPDSSRALLGWPTPLVPALYGLV	LVVGLPANGALWVLTQAPRLP	STMLLNLAT	120		
DB	61	LELPDSSRALLGWPTPLVPALYGLV	LVVGLPANGALWVLTQAPRLP	STMLLNLAT	120		

DT 01-MAY-1999 (T-EMBLrel. 10, Last annotation update)
DE G-PROTEIN COUPLED P2Y NUCLEOTIDE RECEPTOR.
GN TPZY.
OS Meleagris gallopavo (Common turkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Meleagridae; Meleagris.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BLOOD;
RX MEDLINE; 98086419.
RA BOYER J.L., WALDO G.L., HARDEN T.K.;
RT "Molecular cloning and expression of an avian G protein-coupled P2Y
receptor."
RL Mol. Pharmacol. 52:928-934(1997).
DR EMBL; AF031897; AAC60339.1; -.
DR PFM; PF00001; 7tm1; 1.
SQ SEQUENCE 374 AA; 42594 MW; 5320428C CRC32;

Query Match 17.7%; Score 353; DB 13; Length 374;
Best Local Similarity 27.0%; Pred. No. 2.3e-19;
Matches 89; Conservative 67; Mismatches 140; Indels 34; Gaps 9;

QY 79 LPALYGLVVLVGLPANGALWLATO-APRLPSTMLMNLATADLLALALPPRIAYHL 137
DB 43 LLLPSYGVFVGLPLNSWAMNIFVSEMRPNATYTFNENLAISDTLYVSLTLYYYA 102

QY 138 RQGRWPFGEACRLATAALYGHMYGVSLLAAVSLDRYLALVHLRRLARALGRRLALGLC 197
DB 103 DRNNWPGKVFCKIVRFELFVANLYSSILFTCTISVHRMGICHPIRSKWKVKHARLIC 162

QY 198 MAWLMAALALPLTLQRTFLRSLDRVCHDALPDAQASHQWPAFTCLALGCLPL 257
DB 163 VGVNLVVTICLIPLNIFVTT--SKDNSTLCHDTTRPE-EFDHYHYSSINALLFGIPF 219

QY 258 LAMLCVCGATLHTL-----AASGRYGH--ALRLTAVVLASAVAFVPSNLLLLHY 308
DB 220 LVIVCYCLMAKRLCKRSFSPRVPYSKRSIKMIIIVLTFACFVFPHTRILYIT 279

QY 309 -----DPSPSAWGNYGAYVPSLALSTLNSCVDPFYIYVSAEPRDKVRAGLFQR----- 359
DB 280 SRYFOADQCLNIINFYTKITRPLASINSLCDPLTFYFMAGDKYGRGLRRGAQRPRVPT 339

QY 359 -----SP-----GTVASKASAEGRSGMT 379
DB 340 SLLALVSPSDSVSGSCNCE--SRGMT 367

RESULT 5
ID O15132 PRELIMINARY; PRT; 370 AA.
AC O15132;
DT 01-JAN-1998 (T-EMBLrel. 05, Created)
DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (T-EMBLrel. 08, Last annotation update)
DE P2Y5-LIKE RECEPTOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97366605.
RA JANSSENS R., BOEYNAEMS J.M., GODART M., COMMUNI D.;
RT "Cloning of a human heptahelical receptor closely related to the P2Y5
receptor."
RL Biochem. Biophys. Res. Commun. 236:106-112(1997).
DR EMBL; AF005419; AAB66322.1; -.
DR PFM; PF00001; 7tm1; 1.
SQ SEQUENCE 370 AA; 41861 MW; A42AB01D CRC32;

Query Match 17.4%; Score 345.5; DB 4; Length 370;
Best Local Similarity 29.2%; Pred. No. 8.5e-19;

Matches 95; Conservative 61; Mismatches 138; Indels 31; Gaps 9;

QY 37 DSTPSILPAPRGYPGQVCANDSTLELPDSSRALLGHWVPTRLVPALYGLVLVGLPANG 96
DB 13 DSNSSLRPR-----LGNATAN--NTCIVDSDSKYKLNGL-----AVYSVVFILGLITNS 58

QY 97 LALWVLATQAP-RLPSTMLMNLATADLLALALPPRIAYHLRQGRWPFGEACRLATAA 155
DB 59 VSLFVFCFRMKMRSETAIFITNLAVSDLLFVCTLPKIFYNF-NRHWPFGLTICKISGTA 117

QY 156 LYGHMYGVSLLAAVSLDRYLALVHLRRLARALGRRLALGLCMAALMAALPLTLQRT 215
DB 118 FLTNIYGSMLFTCISVDRLAIVYFPRSTITRNSAIVCAGWILVLSGGISASLES 177

QY 216 QTFRLARSLRVCHDALPDAQASHQWPAFTCLALGCLPLLAMLLCYGATLH----- 270
DB 178 TT--NVNNATTTCFEGLSKRVKTYLSKITFIEVGVFIPLILNVSCSVLRLTKPA 235

QY 270 TLAASGRYGHALRLTAVVLASAVAFVPSNLLLLHYSDPSAWGNLY-----GAYV 323
DB 236 TISQIGTNKKVKLMTIVHMAVFCVPYNSVLF-LYALVRSQAITNCFLEFAKIMYP 294

QY 324 PSLALSTLNSCVDPFYIYVSAEFR 348
DB 295 IYLCIATLNCDDPFYIYVTSLESFQ 319

RESULT 6
ID O97666 PRELIMINARY; PRT; 380 AA.
AC O97666;
DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (T-EMBLrel. 10, Last annotation update)
DE PUTATIVE CHEMOKINE RECEPTOR.
GN APJ.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Cercopitheidae; Cercopitheciae;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN;
RA MARGULIES B.J., HAUER D.A., CLEMENTS J.E.;
RT "Identification and characterization of thirteen rhesus macaque
chemokine receptors and chemokine receptor homologues."
RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
CC 1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC 1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL; AF100206; AAC74404.1; -.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR; 1.
KW Receptor; G-protein coupled receptor; Transmembrane; Glycoprotein.
SQ SEQUENCE 380 AA; 42606 MW; D209A573 CRC32;

Query Match 17.3%; Score 344; DB 6; Length 380;
Best Local Similarity 29.5%; Pred. No. 1.1e-18;
Matches 97; Conservative 59; Mismatches 141; Indels 32; Gaps 9;

QY 79 LPALYGLVVLVGLPANGALWL--ATQAPRLPSTMLMNLATADLLALALPPRIAYH 136
DB 30 LIPAIYMLVFLLTGTTGNGLVLTFRSSKRKRSADIFTASLVAQTFFVTIPLWATYT 89

QY 137 LRQGRWPFGEACRLATAALYGHMYGVSLLAAVSLDRYLALVHLRRLARALGRRLALGL 196
DB 90 YRDYDWPFGFTFCKLSYLLFVNMYASVCLTGLSFDRLAIVRPVANRLRVRSGAVA 149

QY 197 CMAALMAALALPLTLQRTFLRSLDRVCHDALPDAQASH---WQPAF-TCIALLG 252
DB 150 TAVLWVLAAALLAMPVNVFTTGDLENTTKQCYMDYSWATVSSDWAWEVGLCVSTTVG 209

QY 253 CFLPLAMLLCYGATLHTLAASGR-----YGHALRLTAVVLASAVAF---FVPSNLLLL 304

Db 210 FVPPTIMLTCTFFTAQIAGHFRKERIEGLRKRRLLSIIIVLVVTFALCWMPYHLVKT 269
Qy 305 LHYSPPSPSAWNLXGAVPSL-----ALSTLNSLQDPFIYIYVSAEPRDKVR----- 353
Db 270 L-YMLGSLHLHPCDFDLFMVFPYCTCISYVNSCLNFIAXFFDPRFRQACTSMLCGG 328
Qy 353 ---AGLFQSPGDTVASKASAEGRSGRMG 378
Db 329 SRCAGTSHSSSGEKSASYSS--GHSQGP 355
RESULT 7
ID 070129 PRELIMINARY; PRT; 345 AA.
AC 070129;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DE 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ANAPHYLATOXIN C5A RECEPTOR.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HARTLEY;
RX MEDLINE; 98236152.
RA FUKUOKA Y., EMBER J.A., YASUI A., HUGLI T.E.;
RT "Cloning and characterization of the guinea pig C5a anaphylatoxin
receptor: interspecies diversity among the C5a receptors.";
RL Int. Immunol. 10:275-283(1998).
DR EMBL; U86103; AAC40074.1; -.
DR PFAM; PF00001; 7tm_1; 1.
SQ SEQUENCE 345 AA; 38592 MW; A47BBFC0 CRC32;

Query Match 17.1%; Score 340.5; DB 11; Length 345;
Best Local Similarity 27.3%; Pred. No. 1.9e-18;
Matches 94; Conservative 62; Mismatches 149; Indels 39; Gaps 6;

Qy 57 DSDTLELPDSSRALLGWVPTRLVPALYG-----LVLVVGPLPANGALWVLATQAP 107
Db 10 DYNSTFLPD-----GFVDNVERLSFGDLVAVVIMVVLVGVPGNALVWVTACEAR 62
Qy 108 RLPSTMLMNLNATADLLALALPPRIAYHLRQGRWPFGEAACRLATAALYGHMYSVLL 167
Db 63 RHINAIWFLNLAADLLSCLAPILLVSTVHLNHWYFGDTACKVLPSSLINMYTISILL 122
Qy 168 AAVSLDRYALVHPLRALRGRRLALGLCHMAWLMMAALPLTLQRTFLARSDRYL 227
Db 123 ATISADRLVLSPIWCQRFQGGCLAWTACGLAWYIALLLSPSPFLYRRTTNEHFSFKY 182
Qy 228 CHDALPLDAQASHWQPAFTCLALLGCFPLMLAMLCYG-ATLHTLAASGRYGHALRLTA 286
Db 183 C--VTDYGRDISKEAVALVLLVGFVPLTLTACYTFLLRTWSRKATSAKTVKVV 240
Qy 287 VVLASAVAFVPSNL-LLLHYSDPSPSANGNLGAYVPSLALSTLNSCDPFIYIYVSA 345
Db 241 AVSSFFFWLPYQVTGILLAWHSPNSATYRNTKALDVCVAFAYINCCINPIIVVWAG 300
Qy 346 EFRDKVRAGL-----FQSPGDTVASKASA 370
Db 301 GFQGRLLKSLPSVLRNVLTRESLDKRHOSFARSTVDTMPQKSES 344

RESULT 8
Q92280
ID Q92280 PRELIMINARY; PRT; 371 AA.
AC Q92280;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE GALANIN RECEPTOR 2.
OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-129/SV;
RX MEDLINE; 99048962.
RA HAWES B., WONG G., WANG S.;
RA HAWES B., WONG G., WANG S.;
RT "The mouse GalR2 galanin receptor: genomic organization, cDNA
cloning, and functional characterization.";
RL J. Neurochem. 71:2252-2259(1998).
DR EMBL; AF077375; AAC95468.1; -.
KW Receptor.
SQ SEQUENCE 371 AA; 40883 MW; 4B6912E5 CRC32;

Query Match 16.9%; Score 336; DB 11; Length 371;
Best Local Similarity 32.5%; Pred. No. 4.4e-18;
Matches 111; Conservative 44; Mismatches 157; Indels 30; Gaps 11;

Qy 56 NDSPTLELPDSSRALLGWVPTRLVPALYGVLVGLPANGALWVLATQAPRLPST-M 113
Db 2 NGSDSQGAEDSSQGGGQWQPEAVLPLFFALPLVGVANGNALVLAFLRGQAVSTNL 61
Qy 114 LLMNLATADLLALALPPRIAYHLRQGRWPFGEAACRLATAALYGHMYSVLLAAVSLD 173
Db 62 FILNLGVADLCFILCCVPFQATITLDDWFGSLCKAVHFLITLTHASSFTLAAVSLD 121
Qy 174 RYLAIVPLRLARALGRRLALGLCHMAWLMMAALPLTLQRTFLARSDRYLCHDALP 233
Db 122 RYLAIRYPLSHRELTPRNALAAIGLIWGLALLSGPYLSYQSOLA--NLTVCHPA-- 178
Qy 234 LDAQASHWQPAFTCLALLGCFPLMLAMLCYGATLHTL-----AASGRRYGHARLT 285
Db 178 WSAPRPWNSCTFCLSYL---LPVLVLSLYARTLTLWRTVDPVAGSSQRAKR-KVT 233
Qy 286 AVVLASAVAF---FVPSNLALLHYSDPSPSANGNLGAYVPSLALSTLNSCDPFIYIY 342
Db 234 RMIVAVVFLCWMHPHALILCVWFGFPLTRAT-YALRILSHLVSVANSVNPVIVAL 292
Qy 343 VSAEFRD---KVRAGLFQSPGDTVASKASAEGRSGRMGTHS 381
Db 293 VSKHFRKGRKICAGLLRRAP-----RRASGRVCILAXGNHS 329

RESULT 9
O35811
ID O35811 PRELIMINARY; PRT; 361 AA.
AC O35811;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE G-PROTEIN COUPLED RECEPTOR.
GN P214.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=LIVER;
RA BOGDANOV Y.D., WILDMAN S., KING B.F., BURNTOCK G.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE; 98421785.
RA WEBB T.E., HENDERSON D., ROBERTS J.A., BARNARD E.A.;
RT "Molecular cloning and characterization of the rat p2y4 receptor.";
RL J. Neurochem. 71:1424-1434(1998).
DR EMBL; Y14705; CAA75007.1; -.
DR EMBL; Y11433; CAA72241.1; -.
DR PFAM; PF00001; 7tm_1; 1.
KW G-protein coupled receptor.


```

189 GRIALGLCWAUWMAALALP-LTLQROFRLARSDRVLC HDALPDAQASHWQPAFTC 247
    ||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
138 TPRNALAAIGLINGISLLFSGPYLSYRQS-QLA--NLTVCHPAW-----SAPRRAMDIC 190
    ||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
248 LALLGCFPLPMLLCLYGATLHTL-----NAGSGRYGHALRLTAVVLASAVAF---F 296
    ||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
191 TTFVSYLLPVLVGLTVARTLYLWRAPDVAAGSGARRAKR-KVTRMILLIIVAALFCLCW 249
    ||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
297 VPSNLLLLHYSDPSAWGNLXGAYPSLSTLNSCVDPFTYYIVSVSAEFRDKVR---A 353
    ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
250 MPHALLTLCVWFQGFPLTRAT-YALRLTSLHVSIVANSVPIVYALVSKHFKGFTICA 308
    ||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
354 GLQRPSPGDIVASKASAEAGSGRGMTHS-SLLQ 385
    ||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
309 GLLGRAPG-----RAGRGVCAAAARGTHSGSVLE 336
    ||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::

RESULT 12
O88626 PRELIMINARY; PRT; 370 AA.
AC O88626;
DT 01-NOV-1998 (TReMBrel. 08, Created)
DT 01-NOV-1998 (TReMBrel. 08, Last sequence update)
DT 01-MAY-1999 (TReMBrel. 10, Last annotation update)
DE GALANIN RECEPTOR TYPE 3.
GN GALR3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
[]]
RN SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE DAWLEY; TISSUE=HYPOTHALAMUS;
RA WATERS S.M., KRAUSE J.E.;
RL "Pat galanin receptor type 3 (Galr3) coding region.";
RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF079844; AAC34590.1; -.
DP PFAM; PF00001; 7tm1.1.
SQ SEQUENCE 370 AA; 40410 MW; 1EBF1E77 CRC32;

```

Query Match 15.9%; Score 317.5; DB 11; Length 370;
Best Local Similarity 33.7%; Pred. No. 1.1e-16;
Matches 104; Conservative 32; Mismatches 126; Indels 47; Gaps 10;

Qy	73	GWPTRLVPALYGLVLVYGLPANGALMWL-----ATQAPRLPSTLLMLNLTADLLLA	126
		: : : : : :	
Db	13	GSVGAVAPVFIALFIFLGVMGNGLVLAIVLQPSPAWQEPRSTDLFILNLAVADLCFI	72
Qy	127	LAIPPRIAYHLRQRHPFGGEACRLATAAIIYGHMYSVLLAAVSDDRYLALVHPLURAR	186
		: :	
Db	73	LCCVPFOAAIYTIDAMLEGAFCVKTVHLLIYTLTWYASFTTAAVSDRYLAVRHPLRSRA	132
Qy	187	L----RGRRLAGLCMAWLMWAALAIAPLTIQRQTFRLASDRDYLCDHALPLDQAASHWOP	243
Db	133	LRTPRNARAAGVL---VWLILAALFSAPYLSYYGTVR-----YCALCVCVA--WED	178
Qy	244	A-----FTCLALLGCFLPLLAMLILCYGATLHTL-----AASGRK--YGHALURLTA	286
		:	
Db	179	ARRALDVATFAGYLLLPVAVSLAIGRTLCTFLWAAVGPAGAAAARRRAYTRAGRAML	238
Qy	287	VVLASAFAFFPSNLIILLHYSDP---SPSAWNLYCAVVPSIALSTLNCSVDPFYYYY	343
		: :	
Db	239	AVAALVALCWGPHALLICFWYGRFAESPAT-----YACLASHCLAYANSLNPVYSLA	294
Qy	344	SAREDKVR	352
Db	295	SRHFRRFR	303

RESULT 13		
093239		
ID 093239	PRELIMINARY;	PRT; 342 AA.
AC 093239;		

DT 01-NOV-1998 (TEMBLrel. 08, Created)
DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TEMBLrel. 10, Last annotation update)
DE CXC CHEMOKINE RECEPTOR-2.
OS Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprininae; Cyprinidae; Cyprinus.
RN [1]
RP SEQUENCE FROM N.A.
RT FUJIKI K., NAKAO M., SHIN D., YANO T.;
RT "cDNA cloning of a carp homologue-2 of mammalian interleukin-8
RT receptors.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
LR EMBL; AB010713; BAA31470.1; -.
RF PFAM; PF00001; 7tm_1; 1.
SQ SEQUENCE 342 AA; 38481 MW; D3370D68 CRC32;

[illegible]

RESULT	14
O60755	
ID	O60755 PRELIMINARY; PRT; 368 AA.
AC	O60755;
DT	01-AUG-1998 (TrEMBLrel. 07, Created)
DT	01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT	01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE	GALANIN 2 RECEPTOR HOMOLOG.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC	Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN	[1]
RN	SEQUENCE FROM N.A.
RP	TISSUE=UTERUS;
RC	RA BENNETT M.M., LESCOE M.K., GALLIOLI P.Z., RAMABHADRAN T.V.;
RL	Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN	[2]
RN	SEQUENCE FROM N.A.
RP	MEDLINE; 98389766.
RX	SMITH K.E., WALKER M.W., ARTYMYSHYN R., BARD J., BOROWSKY B.,
RA	TAMM J.A., YAO W.-J., VAYSSE P.J.-J., BRANCHEK T.A., GERALD C.,
RA	JONES K.A.;
RA	"Cloned human and rat galanin GALR3 receptors. Pharmacology and
RT	activation of G-protein inwardly rectifying K+ channels.";
RL	J. Biol. Chem. 273:23321-23326(1998).

This Page Blank (uspt)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 1999, 19:54:39 ; Search time 1202.25 Seconds
(without alignments)
8031.234 Million cell updates/sec

Title: US-09-371-333-1
Perfect score: 4895
Sequence: 1 CTCGCCAGCGCTGGCTGGCA.....AAAAAAAAAAAAAAAAAAAA 4895

Scoring table: IDENTITY_NUC
Searched: 2546578 seqs, 986266752 residues

Database : EST.*

- 1: em_est1.*
- 2: em_est2.*
- 3: em_est3.*
- 4: em_est4.*
- 5: em_est5.*
- 6: em_est6.*
- 7: em_est7.*
- 8: em_est8.*
- 9: em_est9.*
- 10: em_est10.*
- 11: em_est11.*
- 12: em_est12.*
- 13: em_est13.*
- 14: em_est14.*
- 15: em_est15.*
- 16: em_est16.*
- 17: em_est17.*
- 18: em_est18.*
- 19: em_est19.*
- 20: gb_est1.*
- 21: gb_est2.*
- 22: gb_est3.*
- 23: gb_est4.*
- 24: gb_est5.*
- 25: gb_est6.*
- 26: gb_est7.*
- 27: gb_est8.*
- 28: gb_est9.*
- 29: gb_est10.*
- 30: gb_est11.*
- 31: gb_est12.*
- 32: gb_est13.*
- 33: gb_est14.*
- 34: gb_est15.*
- 35: gb_est16.*
- 36: gb_est17.*
- 37: gb_est18.*
- 38: gb_est19.*
- 39: gb_est20.*
- 40: gb_est21.*
- 41: gb_est22.*
- 42: gb_est23.*
- 43: gb_est24.*
- 44: gb_est25.*
- 45: gb_est26.*
- 46: gb_est27.*
- 47: gb_est28.*
- 48: gb_est29.*
- 49: gb_est30.*
- 50: gb_est31.*
- 51: gb_est32.*
- 52: em_est20.*
- 53: em_est21.*

54: em_est22.*
55: em_est23.*
56: em_est24.*
57: em_est25.*
58: em_est26.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	553.4	11.3	570	51	AI739295	AI739295 w130b03.x
C 2	530	10.8	544	45	AI344017	AI344017 tc01e02.x
C 3	291.8	6.0	611	53	HS0005397	AI040921 Homo sapi
C 4	282	5.8	985	54	HS0009189	AI044339 Homo sapi
C 5	278.6	5.7	542	28	AA081138	AA081138 zn34h07.s
C 6	274	5.6	540	49	AL048626	AL048626 DKF2p564F
C 7	261.2	5.3	529	37	AA680243	AA680243 ac86a01.s
C 8	259.6	5.3	504	35	AA577824	AA577824 nn24e06.s
C 9	259.2	5.3	487	37	AA708108	AA708108 zg06b11.s
C 10	257.2	5.3	457	31	AA284247	AA284247 zc65d01.T
C 11	255.2	5.2	496	41	AI038990	AI038990 ox24f02.x
C 12	255.2	5.2	487	50	AI679002	AI679002 tu60g01.x
C 13	253	5.2	490	49	AI653525	AI653525 tq95e03.x
C 14	252.8	5.2	523	41	AI054414	AI054414 q176h03.x
C 15	252	5.1	569	49	AL048969	AL048969 DKF2p434M
C 16	251.6	5.1	521	44	AI302688	AI302688 qn17a10.x
C 17	247.6	5.1	500	51	AI732911	AI732911 oet2lg10.x
C 18	247.2	5.1	535	48	AI608771	AI608771 tw94f06.x
C 19	240.8	4.9	660	49	AL047602	AL047602 DKF2p586G
C 20	239.6	4.9	630	46	AI457389	AI457389 t173c12.x
C 21	239	4.9	494	34	AA486896	AA486896 ab17h07.s
C 22	235.6	4.8	711	54	HS0009190	AL044340 Homo sapi
C 23	233.6	4.8	524	21	T63408	T63408 yc23e07.s1
C 24	233.6	4.8	443	47	AI468006	AI468006 tj84f09.x
C 25	231.6	4.7	552	35	AA594229	AA594229 nn29h05.s
C 26	231.2	4.7	543	36	AA608741	AA608741 ae56g03.s
C 27	230.8	4.7	441	51	AI732120	AI732120 ac86a01.x
C 28	230.8	4.7	441	51	AI732180	AI732180 ac85g01.x
C 29	230.8	4.7	646	54	HS0008595	AI043745 Homo sapi
C 30	230.8	4.7	591	54	HS0011470	AI046620 Homo sapi
C 31	230.2	4.7	732	53	HS0007755	AI042905 Homo sapi
C 32	229.8	4.7	474	45	AI354847	AI354847 qt76a04.x
C 33	229.4	4.7	430	53	HS0005489	AI041013 Homo sapi
C 34	229	4.7	563	36	AA601355	AA601355 nol8b04.s
C 35	228.2	4.7	554	34	AA524604	AA524604 nh34c01.s
C 36	227.4	4.6	546	48	AI567106	AI567106 tq85c11.x
C 37	227	4.6	492	36	AA614180	AA614180 no94g11.s
C 38	227	4.6	783	54	HS0010804	AI045954 Homo sapi
C 39	226.4	4.6	440	36	AA618412	AA618412 nn27d01.s
C 40	226.4	4.6	478	50	AI683547	AI683547 tx67g09.x
C 41	226.2	4.6	435	34	AA487119	AA487119 ab19a06.s
C 42	226.2	4.6	474	34	AA493641	AA493641 nh03b10.s
C 43	226.2	4.6	492	34	AA534064	AA534064 nj97c05.s
C 44	226.2	4.6	604	35	AA573207	AA573207 nj42f12.s
C 45	225.6	4.6	455	40	AA904211	AA904211 od88e02.s

ALIGNMENTS

RESULT 1
AI739295/c
LOCUS AI739295 570 bp mRNA EST 18-JUN-1999
DEFINITION w130b03.x1 NCI_CGAP_Col6 Homo sapiens cDNA clone IMAGE:2391725 3', similar to TR:076067 076067 PROTEASE-ACTIVATED RECEPTOR 4. ; mRNA sequence.
ACCESSION AI739295
NID 95101276

VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
AI739295.1	GI:5101276	EST.	human.					
			Homo sapiens					
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;					
			Eutheria; Primates; Catarrhini; Hominidae; Homo.					
			1 (bases 1 to 570)					
			NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .					
			National Cancer Institute, Cancer Genome Anatomy Project (CGAP),					
			Tumor Gene Index					
			Unpublished (1997)					
			On Jun 5, 1998 this sequence version replaced gi:3189581.					
			Contact: Robert Strausberg, Ph.D.					
			Tel: (301) 496-1550					
			Email: Robert.Strausberg@nih.gov					
			Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,					
			M.D., Ph.D.					
			CDNA Library Preparation: M. Bento Soares, Ph.D.					
			CDNA Library Arrayed by: Greg Lennon, Ph.D.					
			DNA Sequencing by: Washington University Genome Sequencing Center					
			Clone distribution: NCI-CGAP clone distribution information can be					
			found through the I.M.A.G.E. Consortium/LLNL at:					
			www-bio.llnl.gov/bbrp/image/image.html					
			Seq primer: -40UP from Gibco					
			High quality sequence stop: 453.					
			Location/Qualifiers					
			1. 570					
			/organism="Homo sapiens"					
			/db_xref="taxon:9606"					
			/clone="IMAGE:2391725"					
			/clone_lib="NCI-CGAP_Col6"					
			/tissue_type="colon tumor, RER+"					
			/lab_host="DH10B"					
			/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a					
			modified polylinker; Site_1: Not I; Site_2: Eco RI;					
			Plasmid DNA from the normalized library NCI-CGAP_Col6					
			prepared, and ss circles were made in vitro. Following HAP					
			purification, this DNA was used as tracer in a subtractive					
			hybridization reaction. The driver was PCR-amplified cDNAs					
			from a pool of 5,000 clones made from the same library					
			(cloneids 1057416-1061255, and 1144584-1145351).					
			Subtraction by Bento Soares and M. Fatima Bonaldo.					
			101 a 192 c 211 g 65 t 1 others					
			BASE COUNT					
			ORIGIN					
			Query Match 11.3%; Score 553.4; DB 51; Length 570;					
			Best Local Similarity 99.0%; Pred. No. 1e-57;					
			Matches 568; Conservative 0; Mismatches 2; Indels 4; Gaps 1;					
			QY 176 ATGTGGGGGACGTCTCTGTGGCCCCCTGGTGGTTCAGGCTGTCTGGCGGCACC 235					
			Db 570 ATGTGGGGGACGTCTCTGTGGCCCCCTGGTGGTTCAGGCTGTCTGGCGGCACC 511					
			QY 236 CAGACCCCCACGCTCTACGACGAGAGCGGGAGCACCAGGAGTGGTATGACAGCAGCGCC 295					
			Db 510 CAGACCCCCACGCTCTACGACGAGAGCGGGAGCACCAGGAGTGGTATGACAGCAGCGCC 455					
			QY 296 TCAATCTGTGCTGCCCCCGCGCTACCCAGGCCAAGTCTGTGCCAATGACAGTGCACACC 355					
			Db 454 TCAATCTGTGCTGCCCCCGCGCTACACAGGCCAAGTCTGTGCCAATGACAGTGCACACC 395					
			QY 356 CTGGAGCTCCGGACAGCTCAGCGGCACATCTCTCTGGGCTGGTGCCACAGGCTGGTG 415					
			Db 394 CTGGAGCTCCGGACAGCTCAGCGGCACATCTCTCTGGGCTGGTGCCACAGGCTGGTG 335					
			QY 416 CCGCCCTCTATGGCTGTGTCCTGTGGTGGTGGGCTCCCGGCCAATGGCTGGCTGTGG 475					
			Db 334 CCGCCCTCTATGGCTGTGTCCTGTGGTGGTGGGCTCCCGGCCAATGGCTGGCTGTGG 275					
			QY 476 GTGCTGCGCCACGAGGACACCTCGGCTGCCTCCACCATGCTGCTGATGAACCTCGCGACT 535					

Db

BASE COUNT
ORIGIN

98 a	182 c	203 g	61 t
------	-------	-------	------

Subtraction by Bento Soares and M. Fatima Ronaldo.

Query Match	10.98;	Score 530;	DB 45;	Length 544;
Best Local Similarity	99.3%;	Pred. No. 6.3e-55;		
Matches 544;	Conservative 0;	Mismatches 0;	Indels 4;	Gaps
Qy	245	AGCGTCTACGACGAGACGGGAGCACC	GGAGGTGGTGATGACACGACGACGCCCTCAATCCTG	304
Db	544	AGCGTCTACGACGAGACGGGAGCACC	GGAGGTGGTGAT---GCAGCCCTCAATCCTG	489
Qy	305	CCTGCCCCCGCGCTTACCAGAGCCAA	GTGTGCCAATGACAGTGCACACCTTGAGAGTCTC	364
Db	488	CCTGCCCCCGCGCTTACCAGAGCCAA	GTGTGCCAATGACAGTGCACACCTTGAGAGTCTC	429
Qy	365	CCGACAGACTACGGGCACTGCTTCTG	GGGTGGTGCCACCAAGGCTGGTGCCCGCCCTC	424
Db	428	CCGACAGACTACGGGCACTGCTTCTG	GGGTGGTGCCACCAAGGCTGGTGCCCGCCCTC	369
Qy	425	TATGGGTGTGCTTGGTGGGGCTGCC	CGCCCAATGGGCTGGCGCTTGGGTGCTGGCC	484
Db	368	TATGGGTGTGCTTGGTGGGGCTGCC	CGCCCAATGGGCTGGCGCTTGGGTGCTGGCC	309
Qy	485	ACGACGAGCACTCGGCTGCCCTTCAC	CACTGCTGTATGAACCTCGGACTGCTGACCTC	544
Db	308	ACGACGACACTCGGCTGCCCTTCCAC	ATGCTGTATGAACCTCGGACTGCTGACCTC	249
Qy	545	CTGCTGGCCCTGGCGCTGCCCGCGG	ATCGCTTACCACCTGGTGGCCACGCGTGGCC	604
Db	248	CTGCTGGCCCTGGCGCTGCCCGCGG	ATCGCTTACCACCTGGTGGCCACGCGTGGCC	189
Qy	605	TTCCGGGAGGCCCTCGCCGCTGGCC	ACGCGCACTATGGTTCACATGTATGGCTCA	664
Db	188	TTCCGGGAGGCCCTCGCCGCTGGCC	ACGCGCACTATGGTTCACATGTATGGCTCA	129
Qy	665	GTGCTGTGCTGGCCCGCGTCAAGCT	CGATCGCTTACCTGGCCCTGGTGACCCGCTGGG	724
Db	128	GTGCTGTGCTGGCCCGCGTCAAGCT	CGATCGCTTACCTGGCCCTGGTGACCCGCTGGG	69
Qy	725	GCCCGCGCCCTGCGTGGCGCGCGCT	CGCCCTTGGACTTCGATGGCTGCTTGGCTCATG	784
Db	68	GCCCGCGCCCTGCGTGGCGCGCGCT	CGCCCTTGGACTTCGATGGCTGCTTGGCTCATG	9
Qy	785	GCGGCGCG	792	
Db	8	GCGGCGCG	1	

```

RESULT      3
HSM005397/C
ID   HSM005397 standard; RNA; EST; 611 BP.
XX
XX AC AL040921;
XX
XX SV AL040921.1
XX
XX NI el399257
XX
XX DT 12-MAR-1999 (Rel. 59, Created)
XX DT 12-MAR-1999 (Rel. 59, Last updated, Version 1)
XX
XX DE Homo sapiens mRNA; EST DKFZp434J0515_s1 (from clone
XX DE DKFZp434J0515)
XX
XX EST; expressed sequence tag.
XX KW
XX OS Homo sapiens (human)
XX OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
XX OC Primates; Catarrhini; Hominidae; Homo.
XX
XX RN {1}

```

RP	1-611	
RA	Bloeker H., Boecher M., Brandt P., Meves W., Gassenhuber J., Wiemann S.;	
RA	*	
RT	Submitted (12-MAR-1999) to the EMBL/GenBank/DDBJ databases.	
RL	MIPS, Am Klopferspitz 18a D-82152 Martinsried, GERMANY	
RL		
XX		
CC	Clone from S. Wiemann, sequenced by GBF within the cDNA sequencing consortium of the German Genome Project	
CC	r1 sequence also available	
CC	This clone is available at the RZPD in Berlin	
CC	Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de	
CC		
XX		
XX		
FH	Key	Location/Qualifiers
FH		
FT	source	1. .611
FT		/db_xref="taxon:9606"
FT		/organism="Homo sapiens"
FT		/clone="DKFZp434J0515"
FT		/clone_lib="434 (synonym: htes3). Vector pSport1; host:
FT		DHI08; sites NotI + SalI"
FT		/dev_stage="adult"
FT		/tissue_type="testis"
XX		
XX		
SO	Sequence 611 BP: 115 A: 168 C: 137 G: 191 T: 0 other:	

[illegible]

```
Db 13 AAAAAAAAAAAAA 1
|||||
RESULT 4
ID HSM009189
XX HSM009189 standard; RNA; EST; 985 BP.
AC AL044339;
XX AL044339.1
XX e1402776
XX 12-MAR-1999 (Rel. 59, Created)
DT 12-MAR-1999 (Rel. 59, Last updated, Version 1)
XX Homo sapiens mRNA; EST DKF2p43A092_r1 (from clone
DE DKF2p43A092)
XX EST; expressed sequence tag.
XX Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
OC Primates; Catarrhini; Hominiidae; Homo.
XX [1]
RN 1-985
RA Ansoerge W., Wirkner U., Meves W., Gassenhuber J., Wiemann S.;
RT Submitted (12-MAR-1999) to the EMBL/GenBank/DBJ databases.
RL MIPS, Am Klopferspitz 18a D-82152 Martinsried, GERMANY
XX Clone from S. Wiemann, sequenced by EMBL within the cDNA
CC sequencing consortium of the German Genome Project
CC sl sequence also available
CC This clone is available at the R2PD in Berlin
CC Please contact the R2PD: Ressourcenzentrum, Heubnerweg 6, 14059
CC Berlin-Charlottenburg, GERMANY; Email: clone@rpd.de
XX Key Location/Qualifiers
FH source
FT 1..985
FT /db_xref="taxon:9606"
FT /organism="Homo sapiens"
FT /clone="DKF2p43A092"
FT /clone_lib="434 (synonym: htes3). Vector pSport1; host
FT DH10B; sites NotI + SalI"
FT /dev_stage="adult"
FT /tissue_type="testis"
XX Sequence 985 BP; 334 A; 218 C; 234 G; 199 T; 0 other;

Query Match 5.8%; Score 282; DB 54; Length 985;
Best Local Similarity 71.8%; Pred. No. 1.5e-25;
Matches 432; Conservative 0; Mismatches 150; Indels 20; Gaps 4;

QY 1908 TTGCTTTAGGCAGGAGTTGGGACAGCTGGGCAACATAGGGGATCCCATCTCTACA 1967
DB 127 TAGTATGTCCCGAGGATTCAGACTATCTCTGGGCAACATAGGAGACTCCATCTCT--- 184
QY 1968 CACAAAAAATTTTAAATGAACAGGACATTTGGCATCGGCTATAGTCCCGGACCTC 2027
DB 184 -ACAAAAAATTTAAATTTGAACAGGACATGTGGCAGATGCTGTGGTCCCGAGCTACAC 242
QY 2028 AGAGGCACAGCGGGAGGATCACTTGAGCTGGGGAGGTGTGGTTCAGTGTAGCTATGA 2087
DB 243 AGGAGCTGAGGTGGAGAGTTCGTTGAGCCAGGAGGTTCGAGGCTGCAGTGCAGCTCA 302
QY 2088 TTGTACCACTGCAGCTCCAGCTGGGCAACAGAGCAAGACTTGTCTCAAAAAATAACAAA 2147
DB 303 TCACACACTGCTCTTCAG-CTTGGCAACAGAGTGGAGGCCCTGTCTCTAAAAAAGAAAAA 361
```

```
QY 2148 CTAAATTA--AAAAAGAAGACGAGAGATAGTGGTGTGGTGGCTCACACTCGCAATCC 2205
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 362 AGAAATTACCTAATATATATAAGAAACAAGCCAGGCACAGTGGCTCACACTATAATCC 421
QY 2206 CAGCACTTTTGAAGCCGAGGTGGGCAGATCATCTGAGGCCAGGAGTTCAAGACAGCCT 2265
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 422 CAGCTCTTTGGAGGCCAAGCGGGTGGACCACTGAGTTCAGGAGTTCAAGACAGCCT 481
QY 2266 GGCTAACATGGTGAATCCTATCTCTACCAAAATACAAAAATAGCCAGCGGTGGTGT 2325
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 482 GGCCAACATAGCAAAACGCGCTCTCTACTAAAAATACAAAAATAGCCGGTGTATGCC 541
QY 2326 GGGCACCTGTACTGGGAGGTGCCCAACCCAGCTACTCTGGGAGGCTGAGTCAGGAGAACG 2385
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 542 ACGTGCCCTGTAAT-----CCCCACTACTCAGGAGGCTGAGCGAGGAATCA 588
QY 2386 CTTGAACCTGGGAGCGGAGGTTCGCGTCAGTGTGAGATGGTGCCACTCTCCAGCCTG 2445
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 589 CTTGAACCCAGGAGGTGGAGGTTCAGTGTGAGCGAGATTTGTGACACTGTACTCCAGCCTG 648
QY 2446 GCGGAAGAGCGGACTCTCTCTCCAAAAAAGAGAGAGGAGGAGGACACAGACACACA 2505
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 649 AGTGACAGCAAGACTCCGCTCTCAAAAAACAAACAAACAAACAAACAAACCCAAAA 708
QY 2506 GA 2507
DB 709 TA 710

RESULT 5
LOCUS AA081138 542 bp mRNA EST 21-OCT-1996
DEFINITION zn34h07.sl Stragene endothelial cell 937223 Homo sapiens CDNA
clone IMAGE:549373 3' similar to contains Alu repetitive element;;
mRNA sequence.
ACCESSION AA081138
NID G1623221
VERSION AA081138.1 GI:1623221
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 542)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chisoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevasaki,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
On Sep 12, 1996 this sequence version replaced gi:1994981.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 442.
Location/Qualifiers
FEATURES
source
1..542
/organism="Homo sapiens"
/db_xref="GDB:3927890"
/db_xref="taxon:9606"
/clone="IMAGE:549373"
```


/clone_lib="Stratagene endothelial cell 937223"
/dev_stage="umbilical vein, 1 passage"
/lab_host="SOLR (kanamycin resistant)"
/note="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dT.
Umbilical vein endothelial cells, passaged once. Average
insert size: 1.0 kb; Uni-ZAP XR vector; -5' adaptor
sequence: 5' GAATTCGACGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTT 3"
BASE COUNT 104 a 161 c 121 g 153 t 3 others
ORIGIN

Query Match 5.78; Score 278.6; DB 28; Length 542;
Best Local Similarity 73.08; Pred. No. 4.7e-25;
Matches 408; Conservative 0; Mismatches 126; Indels 25; Gaps 3;

QY 1885 CTCTGGGAGCCAGATGGAGGATGCTTGAAGCCAGGAGTTGGACCACTGGGCAA 1944
Db 541 CTCAGAGCCNAGTGGAGGATCACTTGGCCAGGAGTTTGTAGTACGCTGGCAA 482
QY 1945 CATAGGGGATCCCTCTCTACACACAAAAATTTTAAATGAACCAAGGATTTGTGCA 2004
Db 481 CTAAGTGAGACTCCATCTCTAC-----AAAAATTAATAATAGCCAGGTGGAGNG 427
QY 2005 TGGCGCTATAGTCCAGCCACTCAAGAGGACAGCGGGGAGGATCACTTGAAGCTGGGAG 2064
Db 426 CATCGCTGTGCTCAGCTATTATGAGGAGGCTTGAAGGAGGAGGATCACTTGAAGCAGGAA 367
QY 2065 GTTGTGGTTCAGTCACTATGATGTACCACTGCCTCCAGCTGGGCAACAGCAGCAAG 2124
Db 366 GTTGAAGTGCATGAGCCGTGATCATGCCACTGCCTCCAGCTGGGCAACAGCAGCAAG 307
QY 2125 ACCTGTCTCAAAATTAACAACTAA-----TTAAAAAAGAGAGCAGAGATAG 2177
Db 306 ACCCTGTCTTAAAAAAGAAAGAAAGAACGACGCTGCTTGAAGAAGCGTACAGTAGG 247
QY 2178 TGGGTGTGGTGCACACCTGCAATCCAGCAGCTTTGGAAGCGCGGAGGTGGGAGATCA 2237
Db 246 TCAGCGGTGTGCTCATGCTGTAATCTTAGCAGCTTTGAGAGGCTGAGGTGGGTAGATCA 187
QY 2238 TCTGAGGCGCAGGTTCAAGACCAAGCTGGCTAACATGTCGAATCTCTATCTACACAA 2297
Db 186 CTTGAGTTCAGGGTTTGAACACAGCTTGGCCACATCGGCAACCCACCTCTACTANA 127
QY 2298 AATCAAAAATTAGCAGCGGTGTGTGGGCACTGTACTTGGGAGGTGCCACCCAGC 2357
Db 126 AATCAAAAATTAGCAGCGGTGTGTGGGCACTGTACTTGGGAGGTGCCACCCAGC 80
QY 2358 TACTGGGAGGCTGAGTCAGGAGATCGCTTGAACCTGGGAGGCGGAGGTGGCGTACG 2417
Db 79 TACTGGGAGGCTGAGGAGGAGATCGCTTGAACCTAGGAGGCGGAGGTGGCGTACG 20
QY 2418 TGAGATGTGCGCACTGCAC 2436
Db 19 CGAGATGCGCCACTGCAC 1

RESULT 6
LOCUS AL048626/c 540 bp mRNA EST 30-APR-1999
DEFINITION DKFP564F1822_r1 564 (synonym: hfb2) Homo sapiens cDNA clone
ACCESSION DKFP564F1822, mRNA sequence.
NID AL048626
VERSION g4727160
KEYWORDS AL048626.1 GI:4727160
SOURCE EST.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 540)
AUTHORS Wambutt, R., Heubner, D., Mewes, W., Gassenhuber, J. and Wiemann, S.

TITLE EST (Wambutt, et al.)
JOURNAL Unpublished (1999)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3188302.
Contact: Wambutt R
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany.
FEATURES
Location/Qualifiers
source 1..540
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFP564F1822"
/clone_lib="564 (synonym: hfb2)"
/tissue_type="brain"
/dev_stage="fetal"
/lab_host="X1-2blue"
/note="Vector: pAMP1; Site_1: NotI; Site_2: SalI"
BASE COUNT 90 a 146 c 125 g 179 t
ORIGIN

Query Match 5.6; Score 274; DB 49; Length 540;
Best Local Similarity 72.88; Pred. No. 1.7e-24;
Matches 396; Conservative 0; Mismatches 140; Indels 8; Gaps 3;

QY 1613 CAACATGGTAAACCCCATCTCTACCAAAAATACAAAAATTTAGTGGGCTTGGTGGCTGG 1672
Db 539 CAACATGGCAAAACCCCTGTATCTACTGAAATACAAAAATTTAGTGGGCTTGGTGGCA 480
QY 1673 CGCTGTAAATCCAGCTACTCAGGAGACTCAGGAGAGGATCGCTTGAACCTGGGAGGC 1732
Db 479 TGCCTGTAAATCCAGCTACTCAGGAGGCTCAGGAGAGGATCGCTTGAACCTCAGGAGGC 420
QY 1733 AGAGGTTCGAGTGAAGCGAGATTCGCGCACTGGCACTCCAGCTCGCTGACAGAGAGCCTG 1792
Db 419 AGGGTTGCATGAGCGGACATCGCACCCTGGACTCCAGCTCGGGCAACAGAGCGAGAC 360
QY 1793 TCTCTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1849
Db 359 TCCGTTTCAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 300
QY 1850 GCCAGGTGCAAGTGCCTCAGCCCTATATCTCAGCAGCTCTGGAGGCGCAAGATGAGGAT 1909
Db 299 GCTGGGTGCGGTGCGCCCTCTCTGTAATCCAGCAGCTTTGGGAGGCTGAGGAGCAG- 241
QY 1910 GCTTGAAGCCAGGAGTTTGGGACAGCCTCGGCAACATAGGGGATCCCATCTCTACACA 1969
Db 240 TCAGGAGTTCAGGAGATTGAGACCATCTGGCTAACACAGCGTGAAACCGAGTCTCTACTAA 181
QY 1970 CAAAAAATTTTAAATGAACCGAGGATTTGGCATCGGCTATAGTCCAGGAGCTCAA 2029
Db 180 AAATACAAA-----AAATGAGCCGGGTATGTGGCGGCGCCTGTAGTCCAGGCTACTCGG 125
QY 2030 GAGGCACAGCGGGGAGGATCACTTGAAGCTTGGAGGTGTGGTTCAGTGAAGCTATGATT 2089
Db 124 GAGGCTAGGCGAGGAGATGCGGTGAACCTTGGGAGCGGAGCTTGCAGAGAGCTGAGATC 65
QY 2090 GTACCACTGCCTCCAGCCCTGGGCAACAGCAGACGCTTGTCTCAAAAAATAAACAACT 2149
Db 64 ACACCACTGCCTCCAGCCCGGCGACAGAGAGAGCTCCGCTCTCAAAAAATAAAAAA 5
QY 2150 AAAA 2153
Db 4 AAAA 1

RESULT 7
LOCUS AA680243/c 529 bp mRNA EST 02-DEC-1997
DEFINITION ac86a01.s1 Stratagene ovary (#937217) Homo sapiens cDNA clone
IMAGE:869448 3', similar to contains Alu repetitive element; contains
element L1 repetitive element ; mRNA sequence.
ACCESSION AA680243

QY	2184	TGTTGGCTCACACCTGCAATCCAGCAGCTTTGGAAGCGCGAGTGGCGAGATCATCTGAG	2243
Db	142	TGTTGGCTCACACATGTAATCCAGCAGCTTTGGGAGTTTGGGCGGTGATCA--CGAG	85
QY	2244	GCCAGGAGTTTCAGACGACCGCTGCTTAACATGTGTAATCTCTCTACCAAAATACA	2303
Db	84	GTCAGGAGTTTCGAGACCGACGACTGGCAACATGTTGTAACACCTTCCCTACTAAAAATACA	25
QY	2304	AAAAATTAGCCAGCGCTGGTGGTGG 2327	
Db	24	AAAAATTAGCCAGCGCTGGTGGTGG 1	
RESULT	8		
AA577824/c			
LOCUS			
DEFINITION			
ACCESSION			
NID			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
BASE COUNT			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
QY	1824	CAATTTTAAAGACGAAAAGTGACGGCGAGTGCGAGTGGCTCACGCCCTATATCTCAGC	1883
Db	496	CATTCCAAATGAAGAAATCTACTACGCGCGCGGTGGTTCAGGCTGTATCCCGC	437
QY	1884	ACTCTGGAGCGCAAGATGAGGATGCTTGAAGCAGGAGTTGGACGACCGCTGGGCA	1943
Db	436	ACTTTGGAGCGCGAGTGGCGGAT-CATGAGTTCAGGATTCGAGACCATCTTGGCTA	378
QY	1944	ACATAGGGGATCCCTCTCTACACAAAAAATTTTTTAATGACGAGGATGTGGC	2003
Db	377	ACACGGTGAACCTCGTCTCTAC-----TAAAAATACAAAAATTAGCGGGCATGTGGC	323
QY	2004	ATGCGCTATAGTCCCGACCATTCAGAGCAGCAGCGGGAGGATCATTTAGCCCTGGGA	2063
Db	322	AGTGCTGTAGTCCCGACCTACTTCGGTAGCTGAGCAGGAGTAATGCGATGAACCCCGGA	263
QY	2064	GGTTGTGGTTCAGTGCAGCTATGATTGTACCTGCACTCCAGCTCCGCTGGGCAACAGAGCA	2123
Db	262	GGCGAGCTTGCAGTGAGCGCAAGATTGTGCGACTGCACTCCAGCTTCGGGCAACAGAGCA	203
QY	2124	GACCTGTCTCAAAAATAAACAACATAAATTTAAAAAAGACGACGAGATAGTGGTG	2183
Db	202	GACTGTGTCTCAAAAAAATAAACAACATAAATTTAAAAAAGATTAATCTACCCGCTTTGGGCGGGG	143

QY	1985	ATGAACCCAGGCATTGTGGCATCGGCTATAGTCCAGGCCACTCAAGAGCCACAGCGCGGA	2044
Dd	444	AGTAGCCGGGCATGGTGGCAGGCACCCTGTAATCCGACTACTCAGAGTGTCAGGCGCAA	385
QY	2045	GGATCACTTG-AGCCCTGGGAGGTGTGGTTGGCAGTGAGCTATGATTGACCACTGCACCT	2103
Dd	384	GAATCTCTTGAACCCAGGAGGAGAAAGTTGCAGTGAGCCGAGATTGGCCCACTGCACCT	325
QY	2104	CAGCCTGGCACAGAGCAGAACCTTGCTCTCAAAATTAACAACATAAATTAATAAAAAG	2163
Dd	324	CAGCCTGGGTGACGACGAGACTGTCTCTCAAAAAAANAANAAGAAGAAG	265
QY	2164	-----AAGACGAGAGATAGTGGGTGTGGGTCAACCTGCAATCCCAGCACTTTGGAA	2218
Dd	264	AAAGAAGAATAAAAAAAAAAGCGGGGTGTGGTNGCTCAGCCTGTGAATCCCAGCACTTTGGGA	205
QY	2219	GGCCGAGGTGGGCAGATCATCTGAGCGCAGGAGTTCAAGACCAGCGCTGGCTAACATGGT	2278
Dd	204	GGCCAGGCAAGTGGATCATCCTGAGGTCSAGAGTTCAAGACCAGCCTGGCCAAATGGT	145
QY	2279	AATCTATCTCTACCAAAATACAAAAATTAGCCAGGCGTGGTGGTGGGCACCTGTACT	2338
Dd	144	AAACCCGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGTGGGAGGCACCTGTAT	85
QY	2339	GGGAGGTGCCCAACCCAGCTACTTGGGGAGGCTGAGTCAGGAGAAATCGCTTGAACCTGGGA	2398
Dd	84	-----CCACGCTACTTGGGAGGCTGAGGCAGGAGANTCGTTGAACCCAGGA	38
QY	2399	GGCGGAGGTTGGGTCAGCTGAGATGGTGCCACTGCA	2435
Dd	37	GACACAGGTTGAAGTGGGCAAGATCATGCCATTGCA	1

RESULT	9
AA708108/c	
LOCUS	
DEFINITION	24-DEC-1997 IMAG05b11.s1 Soares pineal gland N3HPG Homo sapiens cDNA clone Z69C=3925A1 3' similar to contains Alu repetitive element;; mRNA sequence.
ACCESSION	AA708108
NID	92718026
VERSION	AA708108.1
KEYWORDS	GI:2718026
SOURCE	EST.
ORGANISM	human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	(bases 1 to 487)
AUTHORS	Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R. WashU-NCI human EST Project Unpublished (1997)
TITLE	On Nov 6, 1997 this sequence version replaced gi:331867.
JOURNAL	
COMMENT	

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m3 fwd. Et from Amersham
High quality sequence stop: 473.

Location/Qualifiers
1. .487
/organism="Homo sapiens"
/db_xref="GDB:129388"
/db_xref="taxon:9606"
/clone="IMAGE:392541"

FEATURES
source

```

/clone_lib="Soares_pineal_gland_N3HPG"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: pineal gland; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site.1: Not I; Site.2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(GT) primer
[5' TGTTACCAATCTGAAGTGGAGCGCCGCCGTTTTTTTTTTTTTTTT
3'], double-stranded cDNA was size selected, ligated to
Eco RI adapters (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pT7T3
vector (Pharmacia). Library constructed by Bento Soares
and M.Fatima Bonaldo."
95 a 131 c 114 g 147 t

```

Query Match	5.3%	Score 259.2	DB 37	Length 487
Best Local Similarity	74.1%	Pred. No. 9.8e-23		
Matches 378	Conservative 0	Mismatches 108	Indels 24	Gaps 3
QY 1625	ACCCATCTCTACCAAAATACAAAATTAAGCTGGGCTTGCTGGCTGGCCCTGTAATCC	1684		
DB				
QY 487	ACCTGTCTCTACTAGAAATACAAAATTAAGCTGGGATGGTGGTGGCTGTAATCC	428		
DB				
QY 1685	CAGTACTACGAGACTGAGCCAGAAGGATCGCTTGAACCTGGGAGGCAGAGGTTCAGT	1744		
DB				
QY 427	TAGTACTCGGGAGCTGAGCAGGAGAATCGCTTGAACAGGAGCAGAGGTTCAGT	368		
DB				
QY 1745	GAGCGAGATGGCCCACTGGACTCCAGCTCGGTCGAGACAGAGCCCTGCTCTAAATTA	1804		
DB				
QY 367	GAGCAAGATTGCACCACTGCTCCAGCGTGGGTGACAGAG-----CAAGAC	320		
DB				
QY 1805	TTAATTAATTAATTAATTCATTTTAAAGAGCAAGAACTGACGGCCAGGTGCAGTGGC	1864		
DB				
QY 319	TCATCTCAAAAAAAAATTAATAATAAAAAAAGGCGCGGTGGTGGC	260		
DB				
QY 1865	TCAGCCTATAATCTCAGCACTCTGGGAGGCCAAGATG-CAGGATTCCTTGAAGCCAGGA	1923		
DB				
QY 259	TCATGTCGTAAATCCCACTCTGGGAGGCCAAGATGAGTGGATCACTTGAGGCCAAGA	200		
DB				
QY 1924	GTTTGGACACAGCTGGGCAACATAGGGGGATCCATCTCTACACACAAAAAATTTT	1983		
DB				
QY 199	GTTCAAGCAGCGCTGCCAACATGTTGAACACCATCTCT-----AATACAAA	151		
DB				
QY 1984	AATGAACAGGCATTGGGCATGGCCTATAGTCCAGCCACTCAAGAGGCACAGGCGG	2043		
DB				
QY 150	ACTTAGCCTGGCATGGGCGACATGCCCTGTAAATCCCAAGCTACTCTGGAGGCTGAGGCAG	91		
DB				
QY 2044	AGGATCACTTGAGCTGGGAGGTTGGTTCGACAGTGCAGCTATGATTGTACCACTGCAC	2103		
DB				
QY 90	AGATCCCTTGAACCTTGAGAGGTGGAGTTGCAGTGAGCCGAGAGTTCCGCCACTGCAC	31		
DB				
QY 2104	CAGCCTGGGCAACAGACCAAGACCTTGTCT	2133		
DB				
QY 30	CAGCCTGGGCGACAGAACAAAGACCCCTGTCT	1		
DB				

RESULT	10
AA284247	
LOCUS	
DEFINITION	AA284247 457 bp mRNA EST 08-NOV-1997 zc65d01.t37 Soares fetal heart NDH19W Homo sapiens CDNA clone IMAGS:J77169 3' similar to contains Alu repetitive element;contains element TARI PIR5 repetitive element ; mRNA sequence.
ACCESSION	AA284247
NID	G1928547
VERSION	AA284247.1 GI:1928547
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 457)
REFERENCE	
AUTHORS	Hallier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le.N., Lennon,G., Marra,M.,

Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
On May 9, 1995 this sequence version replaced gl:802598.

TITLE
JOURNAL
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Putative full length read

The vector to vector length is 466
Insert Length: 503 Std Error: 0.00
Seq primer: primer name ambiguous
High quality sequence stop: 431.

FEATURES

source

1. 457
/organism="Homo sapiens"
/db_xref="GDB:1261371"
/db_xref="taxon:9606"
/clone="IMAGE:327169"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: p7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGCGCATCTTTTCTTTTCTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHL19W."

BASE COUNT 149 a 98 c 123 g 87 t

ORIGIN

Query Match 5.3%; Score 257.2; DB 31; Length 457;
Best Local Similarity 75.8%; Pred. No. 1.7e-22;
Matches 350; Conservative 0; Mismatches 98; Indels 14; Gaps 2;

QY 2022 CCACCTCAGAGGCACAGCGGGAGGATCACTTGAGCCTGGGAGGTTGTGGTTGCAGTGAG 2081

Db 1 CTACTCAGGAGGCTGAGCGAGGAGATCACTTCAACTGGGAGCGGAGTTGCAGTGAG 60

QY 2082 CTATGATTGATCCACTGCAGCTCAGCCTGGGCAACAGACAGACCTTGTCTCAAAAATA 2141

Db 61 TCGAGATCATGACTGCAGCTCAGCCTAGGTAACAGACAGACCTCCGTCTCAAAAAGA 120

QY 2142 AACAACTAAATTAATAAAGAGACGAGAGATAGTGGTGTGGCTCACACCTGCA 2201

Db 121 AA-AAAAAAGGGATGTTTAAATCTGGCGGCGATGATGGCTCACACTGTA 179

QY 2202 ATCCAGCACTTTGGAAGCCGAGGTGGGAGATCATCTGAGCCAGGAGTTCAAGACCA 2261

Db 180 ATCCAGCACTTTGGGAGGCTGAGCGGGCTGATCACTTGAAGTTCAGGAGTTCGAGACCA 239

QY 2262 GCTGCTAACATGGTGAATCTTCTATCTACCAAAAATACAAAATAGCCAGGGCTGG 2321

Db 240 GCTGCTAACATGGTGAATCTTCTATCTTACTAAAAATACAAAATAGCCAGGTTGG 299

QY 2322 TGGTGGGCACCTTGTACTTGGGGAGGTGCCACCCAGCTACTTGGGGAGGCTGAGTCAGGAGA 2381

Db 300 TGGCAGACACCTGTAT-----CCAGCTACTTGGAAAGGCTGAGGCAGGAGA 346

QY 2382 ATCGCTTGAACCTGGGAGCGAGGTTGGGTACGTGAGATGGTCCACTGCACCTCAG 2441
Db 347 ATCACTTGAACCAAGAGGTGGAGGTTGCAGTGAAGTGTGCACCTGCACTTCAG 406

QY 2442 CCGTGGCGGAAGAGCGACTCTGTCTCCAAAAAAGAGAAGA 2483

Db 407 CCTGGCAACAGAGTGAGACTCCATTTCAAAAAAAGAAAA 448

RESULT 11

AI038990/c

LOCUS

DEFINITION

Ox24f02.x1 Soares_total_fetus_Nb2HF8.9w Homo sapiens cDNA clone

IMAGE:1657275 3' similar to contains Alu repetitive element;; mRNA

sequence.

ACCESSION

AI038990

NID

G3278184

VERSION

AI038990.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE

1 (bases 1 to 496)

AUTHORS

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

On Jan 14, 1998 this sequence version replaced gl:1796960.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 915 Std Error: 0.00

Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 473.

FEATURES

Location/Qualifiers

1. 496

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1657275"

/clone_lib="Soares_total_fetus_Nb2HF8.9w"

/dev_stage="8-9 weeks"

/lab_host="DH10B"

/note="Vector: p7T3D-pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA

was prepared from mRNA obtained from pooled 8-9 week

(total) fetus material with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTGGGAGCGCGCTTAATTTTTTTTTTTT 3'].

Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified p7T3 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 86 a 139 c 129 g 142 t

ORIGIN

Query Match 5.2%; Score 255.2; DB 41; Length 496;
Best Local Similarity 77.7%; Pred. No. 2.9e-22;
Matches 390; Conservative 0; Mismatches 93; Indels 19; Gaps 6;

QY 1974 AAAATTTTATGACAGGCACTTGTCATGCGCCCTATAGTCCAGCCACTCAAGAGG 2033

Db 486 AAAATACAAAATAGCAGGCACTGTCGACACACCTGTATATCCAGCTACTGGAGAGG 427

QY 2034 CACAGCGGGAGGATCACTTGAGCCTGGGAGGTTGTGGTTCAGTGAGCTATGATTGAC 2093

Db 426 CTGAGCGAGGAGAAATCGCTTGAACCCGGAGGCGGAGGTTGCCGTGAGCTGAGATCGCGC 367

QY 2094 CACTGCATCCAGCCTGGGCAAC-AGACGAGACCTTGTCTCAAAAATAAACAATAA 2152

```

|||||
Db 366 CACTGCACTCAGCGCTGGGCAACAGCGGAACTCTGTCTCAAAAAAATTTTATTTT 307
QY 2153 ATTA-AAAAAAGACGAGAGATAGTGGGTGGTGGCTCACACCTGCAATCCCGACAC 2211
Db 306 AGTATATATTTTAAAGACATTTGGCGGTGGTGGCTCACGCTGTATCCCGACAC 247
QY 2212 TTTGGAAGGCGGAGTGGGCGAGATCATCTGAGGCGGAGGTTTCAAGACCGCTGGCTAA 2271
Db 246 TTTGGAGGCGGAGTGGGCGAGATCA--CGAGGTGAGGAGATGAGACCATCTTGGCCAA 189
QY 2272 CATGTGAAATCTCTCTTACCAAAAAATACAAAAATAGCCAGCGCTGGTGGGCGAC 2331
Db 188 CACGGTAAACCCCGTCTCTCTAAATATAC-AAATAGCCAGGATGTTGGGCGACAGC 130
QY 2332 CTGTACTGGGAGTGGCCACCCAGCTACTGGGAGGCTGAGTGAGGAGATCGCTTGA 2391
Db 129 CTGTAAT-----CCAGCTACTGGAGAGGCTGAGGCGAGGAGATCGCTTGA 83
QY 2392 CCTGGGAGGCGGAGTGGGCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2451
Db 82 CCCGGGAGGCGGAGTGGGCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 23
QY 2451 AAGAGCGACTCTGTCTCAAAA 2472
Db 22 AAGAGCGAACTCTGTCTCAA 1

```

```

RESULT 12
AI679002/c
LOCUS
DEFINITION
t160901.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2255472 3'
similar to contains Alu repetitive element; contains element TARI
repetitive element ;, mRNA sequence.
ACCESSION
AI679002
NID
94889184
VERSION
AI679002.1 GI:4889184
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 487)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
On Mar 10, 1998 this sequence version replaced gi:2948847.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 422.
Location/Qualifiers
1. .487
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="788H12; 14q24.3"
/clone="IMAGE:2255472"
/tissue_type="poorly differentiated adenocarcinoma with
signed ring cell features"
/lab_host="DH10B"

```

```

FEATURES
source
1. .487
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="788H12; 14q24.3"
/clone="IMAGE:2255472"
/tissue_type="poorly differentiated adenocarcinoma with
signed ring cell features"
/lab_host="DH10B"

```

```

/note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.69 kb. Life technologies catalog #:
11549-011"
BASE COUNT 104 a 139 c 110 g 134 t
ORIGIN
Query Match 5.2%; Score 255.2; DB 50; Length 487;
Best Local Similarity 73.0%; Pred. No. 2.9e-22;
Matches 375; Conservative 0; Mismatches 98; Indels 41; Gaps 2;
QY 1954 ATCCCATCTCTACACAAAAAATTTTAAATGAACCCAGGCTTGGGATGGGCTAT 2013
Db 476 ATGAGACCCCTTCTCTACAGAAAATTTAAAAATCAGCCAGCATGTGTATGTGCTGT 417
QY 2014 AGTCCCGACCCACTCAAGAGGCGGAGGAGTCACTTGAAGCTGGGAGGTTGTGGTT 2073
Db 416 AGTCCCGAGTTACTCAGAGGCTGAGGCGGAGGATTTGCTTGAGCCAGGAGCTCAAGGCT 357
QY 2074 GCAGTGAGCTATGATGTACCTCCAGCTGAGGCTGGGCAAGAGAGCTTGTCT 2133
Db 356 ACAGTGAGCTATGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 297
QY 2134 CAAAAATAAACAACATAAAATTTAAAAAAGAACGAGAGATAGTGGTGTGGGCTCA 2193
Db 296 TTAATAAATAAAGCA-----GGTGTGGGCTCA 265
QY 2194 CACTGCAATCCAGCACTTTTGAAGCCGAGGTGGGAGCATCATCTGAGGCCAGGAGTT 2253
Db 264 TGCTGTATCCAGCATTTTGGAGGCTGAGGCAAGTGTGATCAGCTGAGGAGTT 205
QY 2254 CAAGACAGCCTGGCTAACATGTGAAATCTATCTACCAAAAAATACAAAAATTAGCC 2313
Db 204 GAAGACAGCCTGACCAACATGTTGAACCGCATCTCTACTAAAAATACAAAAATTAGCT 145
QY 2314 AGCGTGGTGGTGGGCACTGTACTGGGAGGTGCCACCCAGCTACTGGGAGGCTGAG 2373
Db 144 GGGCATGGTGGCAGCATCTGTAAT-----CCAGCTACTTGGGAGGCTGAG 98
QY 2374 TCAGGAGAACTCGTTGAACCTGGGAGGCGGAGGTGGGTGAGTGTGGTGGTGGTGGT 2433
Db 97 GCAGGAGACTGTGTAACCCGCGGAGGCGGAGGTGAGTGTGAGCCGAGATCAAGCCATTG 38
QY 2434 CACTCCAGCTGGGCGGAGAGAGGAGCTCTCTCTC 2467
Db 37 CACTCCAGCTGGGTGACAGAGTGAGACTCCATC 4
RESULT 13
AI653525/c
LOCUS
DEFINITION
tq95e03.x1 NCI_CGAP_Ov23 Homo sapiens cDNA clone IMAGE:2216572 3'
similar to contains Alu repetitive element;; mRNA sequence.
ACCESSION
AI653525
NID
94737504
VERSION
AI653525.1 GI:4737504
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 490)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
On May 18, 1998 this sequence version replaced gi:3137414.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

```

```

BASE COUNT      95 a      145 c      122 g      161 t
ORIGIN
non-amplified.

Query Match          5.2%; Score 252.8; DB 41; Length 523;
Best Local Similarity 73.9%; Pred. No. 5.5e-22;
Matches 390; Conservative 0; Mismatches 102; Indels 36; Gaps 4;

QY 1586 AGCCACGAGGTTCACACACCAGCCTGAGCAACATGGTAAAAACCCCATCTCTACCAAAAAATA 1645
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 523 AGGTCAGGAGTTCAGAGACCAGCCTGGCCCAACACAGCGTGAACACCACGTCTCTACT-AAAAATA 465
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1646 CAAAAAATAGCTGGGCTTGSTGCTGGCGGCTGTAAATCCAGCTACTACAGGAGACTGAGG 1705
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464 CAAAAAATAGCTGGGTGTGGTGTGGCTGTGTAATCCAGCTACTACAGGAGGCTGAGG 405
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1706 CAGAAGGATCGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATTGCGGCACCTGG 1765
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 404 CAGAAGAATCACITGAACCCGGGAGGTGGAGTTGCAGTGAAGCCAGATCAGCCCACTGT 345
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1766 ACTCCAGCCTCGGTGCACAGAGAGCCTCTCTCTAAATTAATTAATTAATTAATTAATCA 1825
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 344 ACTCCAGCCTCGGCGACAGAGTG-----AGACTCCA 314
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1826 ATTTTAAAAAGACAAAAGTGCAGCGCCAGGTGCAAGTGGCTCACGCCCTATAATCTCAGGCAC 1885
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 313 TCTTAAAGAAAAAAGAGCCAGGTACAGTGGCTCATGCTGTATATCCAGCAC 254
Qy 1886 TCTGGAGGCCAA-GATGGAGGATTGCTTGAAGCCAGGAGTTTGGGACCAAGCTGGGCAA 1944
Db 253 TTTGGAGGTCAAGGAGGTGATCAGCTGAGTGCATGAGTTTAAAGACCAAGCTGGGCAA 194
Qy 1945 CATAGGGGATCCCATCTCTACACACAAAAAATTTTATGACACAGGCAATGTGGCA 2004
Db 193 CATGGTAGACCCCATCTCTACTAAAAACAA-----AAATTACCCAGGCAATGTGGCA 139
Qy 2005 TGGCCTATAGTCCAGCCACTCAAGAGGCACAGCGGGAGGATCATTGAGCCTGGGAG 2064
Db 138 GGCACCTGTATCTCAGCTACTTGGAGGTGAGGAGGAGGAGGAGTCTGGAAACCCGGGAG 79
Qy 2065 GTTGTGGTTCAGTGAGCTATGATGTACCACTGCATCCAGCCTGGG 2112
Db 78 GTGGAGGTTCAGTGAGCTTATGATGTGCCATTGCCTACACTACAGCCAGG 31

RESULT 15
AL048969 569 bp mRNA EST 30-APR-1999
LOCUS DKF2p434M1318_r1 434 (synonym: htes3) Homo sapiens cDNA clone
DEFINITION DKF2p434M1318, mRNA sequence.
ACCESSION AL048969
NID 94728278
VERSION AL048969.1 GI:4728278
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 569)
AUTHORS Ottenwaelder, B., Obermaier, B., Mewes, W., Gassenhuber, J. and
Wiemann, S.
TITLE EST (Ottenwaelder, et al.)
JOURNAL Unpublished (1999)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189429.

Contact: Ottenwaelder B
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany.
FEATURES
source
1..569
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="8; 11q13"
/clone="DKF2p434M1318"
/clone_lib="434 (synonym: htes3)"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT 177 a 134 c 157 g 101 t
ORIGIN

Query Match 5.1%; Score 252; DB 49; Length 569;
Best Local Similarity 71.4%; Pred. No. 6.7e-22;
Matches 434; Conservative 0; Mismatches 130; Indels 44; Gaps 6;

Qy 1549 CCAGCATTAAAGGCCAAGCGGATGATCATTGAGCCAGGAGTTCAACACCAAGCC 1608
Db 1 CCAGCATTGGGAGGCCGAGGTGGATTAAGTACGTGGGCAATTTGAGACCAAGCC 60
Qy 1609 TGAGCAACATGTAAACCCCATCTCTACCAAAATACAAAAA-TTAGCTGGGCTGGTG 1667
Db 61 TGGCAAGATGTTGAACCCCGTCTCTACTAAATAGAAAAAGTTAGCTGGGCTGGTG 120
Qy 1668 GCTGGCGCTGTAATCCAGCTACTTCAGGAGCTGAGGAGGATCGCTTGAACCTGG 1727
Db 121 GCGGTCACTGTATATCCAGCTACTTGGGAGGCTGAGGAGGAGTATCGCTTGAACCCAG 180

```

```

Qy 1728 GAGCGACAGGTTGAGTGAAGCCAGGATTGGCCACTGGACTCCAGCTCGTGCACAGAGA 1787
Db 181 GAGCGACAGGTTGAGTGAAGCCAGGATTGGCCACTGGACTCCAGCTCGTGCACAGAGA 236
Qy 1788 GCCTGTCTCTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1847
Db 236 -----AAGAGCAAAACTCCATCTCAAAAAA----- 268
Qy 1848 CGGCAGGTGAGTGGCTCAGCGCTATATATCTCAGCACTCTGGGAGGCCAA-GATGGAGG 1906
Db 268 --AAACTTTGCGGGCTTCAGCGGTGTAATCCAGCACTTTGGGAAGCCCAAGGCGGCGG 325
Qy 1907 ATTGCTTGAAGCCAGGAGTTTGGGACCAAGCTGGGCAACATAGGGGATCCCATCTCTAC 1966
Db 326 ATCACTTGAAGTCAAGGAGTTCCAGACCAACCTGGCCCAACATGTTGAAACCTGTCTCTAT 385
Qy 1967 ACACAAAAAATTTTAAATGAACCCAGGCAATTTGGGCATGCGCTATATAGTCCCAAGCCACT 2026
Db 386 AAAAAATACAA-----AGTTAGCCAGGTGTGGGACCGGCTGTATATCCAGCTACT 439
Qy 2027 CAAGAGCCACAGCGGAGGATCATTGAGCCTGGGAGGTTGTGGTTCAGTGAAGCTATG 2086
Db 440 TGGGAGGCTGAGGCGAGGAGAACCACTTTGAACCGGGAGGCGGAGGTTGCAGTGAAGCTGAG 499
Qy 2087 ATTGTACCACTGCCTCCAGCCTGGGCAAC-AGAGCAAGACCTTGTCTCAAAAAATAACA 2145
Db 500 ATCACTACCACTGCCTCCAGCCTGGGTCGACAAAGAGCAAAACTTCATCTCAAAAAA 559
Qy 2146 AACTAAAA 2153
Db 560 AAAAAAA 567

```

Search completed: October 28, 1999, 21:00:57
Job time: 3978 sec

This Page Blank (uspto)

OM nucleic - nucleic search, using sw model
Run on: October 29, 1999, 00:27:33 ; Search time 164.19 Seconds
(without alignments)
7458.982 Million cell updates/sec
Gencore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

Title: US-09-371-333-1
Perfect score: 4895
Sequence: 1 CCCCCAGGCTGGCTGGCA.....AAAAAAAAAAAAAAAAAAAA 4895

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	642	13.1	2588	1 V40372	Mouse G-protein co
2	618.2	12.6	1080	1 V40373	Mouse G-protein co
3	458.6	9.4	16891	1 X37084	MEFV gene sequence
4	448	9.2	1470	1 Q47355	Myotonic dystrophy
5	420.2	8.6	14796	1 V27941	Survivin gene. Mod
6	402	8.2	32367	1 V35620	Human SHOX (short
7	381.2	7.8	53577	1 T18551	Human polycystic k
8	381.2	7.8	53577	1 T94108	Human PKD1 locus b
9	381.2	7.8	53577	1 T94101	Human PKD1 gene. H
10	366.6	7.5	6901	1 T78853	Human lecithin-cho
11	349	7.1	32367	1 V35620	SHOX gene prelinin
12	345.8	7.1	15577	1 V35616	SHOX gene prelinin
13	339.4	6.9	5543	1 T75284	Nucleotide sequenc
14	336.4	6.9	8342	1 Q75209	ALL-1 (acute lymph
15	331.2	6.8	7146	1 V38933	Nucleic acid seque
16	324.6	6.6	8391	1 T16333	MLL gene 8.3 kb fr
17	323	6.6	8392	1 Q53478	MLL gene 8.3 kb Ba
18	321.6	6.6	13747	1 Q47342	Myotonic dystrophy
19	314.4	6.4	22481	1 T11658	Human kidney amino
20	310.4	6.3	45546	1 X23320	GDP-Fuc:beta-D-gal
21	306.6	6.3	8174	1 Q13332	DNA encoding a gly
22	306.6	6.3	8174	1 Q56908	Human alpha(1,2)-f
23	306.6	6.3	8174	1 T61677	Human alpha(1,2)-f
24	306.6	6.3	3373	1 T76768	Genomic clone enco
25	303.8	6.2	10475	1 N80616	Sequence of human
26	303.8	6.2	10475	1 N80643	SHOX gene exon Vb
27	303.4	6.2	625	1 V35615	Human transcriptio
28	303.4	6.2	1354	1 V35619	FMF associated pro
29	303	6.2	3512	1 X37085	Human Cdn-2 DNA. N
30	302.4	6.2	6511	1 Q95493	Human flavin-conta
31	301.6	6.2	25464	1 V57274	Hereditary haemoch
32	301.6	6.2	23736	1 V57503	DNA encoding macro
33	296.6	6.1	2923	1 T76529	Coding sequence fo
34	296.6	6.1	2923	1 T99233	Macrophage derived
35	296.6	6.1	2923	1 X32817	BRCA1, human breas
36	295.6	6.0	24026	1 T18325	Mutated BRCA1 geno
37	295.6	6.0	24026	1 T17455	Mutated BRCA1 geno
38	295.6	6.0	24026	1 T17512	Mutated BRCA1 geno
39	295.6	6.0	24026	1 T17513	Mutated BRCA1 geno
40	295.6	6.0	24026	1 T17514	Mutated BRCA1 geno
41	295.6	6.0	24025	1 T17515	Mutated BRCA1 geno
42	295.6	6.0	24026	1 T17516	Mutated BRCA1 geno
43	295.6	6.0	24026	1 T17517	Mutated BRCA1 geno

RESULT 1
V40372
ID V40372 standard; cDNA; 2588 BP.
AC V40372;
DT 16-OCT-1998 (first entry)
DE Mouse G-protein coupled receptor 69A08 #1 encoding cDNA.
KW Mouse; CKDLR201.1; chemokine; G-protein coupled receptor; GPCR;
7 transmembrane receptor; inflammation; asthma; antiviral;
KW abnormal cell proliferation; regeneration; degeneration; atrophy; ss.
OS Mus sp.
FH Key Location/Qualifiers
FT 1..1083
FT /*tag= a
FT /product= "G-protein coupled receptor 69A08"
FT misc_feature 158
FT /*tag= b
FT /note= "Residues 158, 159 and 276 are probably absent, changing the reading frame between those positions; sequence are provided in V40373 and W69598"
FT FT
FT PN WO9831810-A2.
PD 23-JUL-1998.
PF 20-JAN-1998; U00218.
PR 21-JAN-1997; US-786624.
PI (SCHE) SCHERING CORP.
PI Capone M, Gorman DM, Hedrick JA, Ruffine CF, Rossi DL,
PI Vicari A, Zlotnik A;
DR WPI; 98-414108/35.
DR P-PSDB; W69597.
PT Pure or recombinant chemokine CKDLR201.1 - useful, e.g. for treating inflammation and as antiviral agents
PS Claim 7; Page 61-64; 77pp; English.
PS The present sequence encodes mouse G-protein coupled receptor 69A08 #1 protein which comprises a plurality of epitopes. Host cells containing vectors comprising a nucleotide sequence encoding the protein are used to produce recombinant protein. Treatment of a cell, particularly neuron, macrophage or lymphocyte, carrying a G-protein coupled receptor (GPCR) or a receptor responsive to CKDLR201.1 chemokine protein, with an antagonist is used to control physiological development, e.g. alteration of calcium ion influx, a chemotactant response, morphology, sequences the CKDLR201.1 protein, or GPCRs, are useful as primers or probes, e.g. for detecting and isolating related sequences and for expressing antigenic peptides. Antibodies (Ab) directed against the CKDLR201.1 protein and GPCRs are used to detect or purify the proteins; diagnostically (e.g. for developmental abnormalities); in screening for potential drugs; to inhibit chemokine/receptor activation; (when coupled to a toxin or radioisotope) for killing specific cells, and to raise anti-idiotypic antibodies. CKDLR201.1 protein and GPCRs and compounds which bind them can be used to treat inflammation, e.g. asthma; as antiviral agents, and to treat abnormal cell proliferation, regeneration, degeneration and atrophy. Therapeutic agents are administered orally, by injection and rectally.
SQ Sequence 2588 BP; 462 A; 782 C; 743 G; 601 T;

Query Match 13.1%; Score 642; DB 1; Length 2588;

Best Local Similarity 74.5%; Pred. No. 1.4e-75;

Matches 876; Conservative 0; Mismatches 290; Indels 10; Gaps 5;

QY 290 AGCCCTCAATCTGCTGCCCGCGGCTACCGAGCAAGTCTGTGCAATGACAGT 349

Db 1198 AAGTCTCTCAGACAAAGCCCTAATCCAGAGGCTACCGGCAAAATCTGTGCAACGACAGT 1257

QY 350 GACACCTGGAGCTCCCGGACAGCTACGGGCACTGTTCTGGCTGGTGGCCACGAGG 409

```
Db 1258 GACACGCTGGAGCTCCCGCCAGCTCTCAAGCACAGCTGCTGCTGGGTGGGTCCCCACGAAG 1317
QY 410 CTGGTCCCGCCCTCTATGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 469
Db 1318 CTGGTACCTGCTCTATGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1377
QY 470 CTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 528
Db 1378 CTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1437
QY 529 CGGAGCTGCTGACCTCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 588
Db 1438 GGCAGTGGCTGATGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1497
QY 589 TGGCCAGCGCTGGCCCTTGGCGGAGCGCCCTGCGCCCTGGCCGCGCCGCGCAGCTCTATGG 648
Db 1498 TGGCCAGCGCTGGCCCTTGGCGGAGCGCCCTGCGCCCTGGCCGCGCCGCGCAGCTCTATGG 1557
QY 649 TCACATGATGCTCAGTCTGCTGCTGGCGCGCTGCTGCTGGCTGGCTGGCTGGCTGGCTGGCT 708
Db 1558 CCACATGATGCTCAGTCTGCTGCTGGCGCGCTGCTGCTGGCTGGCTGGCTGGCTGGCTGGCT 1617
QY 709 GGTGACCGCTGGCGGCGCGCCCTGCGCCCTGGCCGCGCCGCGCAGCTCTATGG 768
Db 1618 GGTGATCCCTTGGCGGCGCGCTGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1677
QY 769 GGTGCTGCTGCTCAGTGGCGCGCCCTGCGCCCTGGCCGCGCCGCGCAGCTCTATGG 828
Db 1678 GGTGCTGCTGCTCAGTGGCGCGCCCTGCGCCCTGGCCGCGCCGCGCAGCTCTATGG 1737
QY 829 CCGGTGGCGCGCTGCTGCTGCTGGCGCGCTGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 888
Db 1738 CC-GATTACTGCTGCTGCTGCTGGCGCGCTGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1796
QY 889 CTCACCTGGCAGCGCCCTTACCTGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 947
Db 1797 CTCACCTGGCAGCGCCCTTACCTGCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1856
QY 948 TGGCCATGCTGCTGCTGCTGCTGCTGGCGCGCCCTGCGCCCTGGCCGCGCCGCGCT 1007
Db 1857 TGGCCATGCTGCTGCTGCTGCTGCTGGCGCGCCCTGCGCCCTGGCCGCGCCGCGCT 1916
QY 1008 AGGGCAGCGCTGAGCTGAGCGAGTGGTGGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1067
Db 1917 ACAGCCTGCTGCTGCTGCTGCTGCTGGCGCGCTGCTGCTGGCTGGCTGGCTGGCTGGCTGGCT 1976
QY 1068 CCAGCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1127
Db 1977 CTAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2036
QY 1128 TCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1187
Db 2037 TCTATGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2096
QY 1188 TCATCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1247
Db 2097 TCATCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2153
QY 1248 GGTGCGGGGAGACGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1307
Db 2154 GGCAGCGGAGGCGCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2213
QY 1308 GCACCCACTCTCTTTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1363
Db 2214 CCATTTGCTCTCTACACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2273
QY 1364 ACAGGGTCCCTTCCCGCCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1423
Db 2274 TGACGTGGGTCTCCCTCTTAACTACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1459
QY 1424 AAATAGGGTGTGTACAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1499
Db 2334 ATATGAGTGTGGTGGACCTTCTATGCTGAGCTGA 2369
```

RESULT 2

```
V40373
ID V40373 standard; cDNA; 1080 BP.
AC V40373;
DT 16-OCT-1998 (first entry)
DE Mouse G-protein coupled receptor 69A08 #2 encoding cDNA.
KW Mouse; CKDLR201.1; chemokine; G-protein coupled receptor; GPCR;
KW 7 transmembrane receptor; inflammation; asthma; antiviral;
KW abnormal cell proliferation; regeneration; degeneration; atrophy; ss.
OS Mus sp.
FH Key
FT CDS
FT Location/Qualifiers
FT 1..1080
FT /*tag= a
FT /product= "G-protein coupled receptor 69A08"
FN W09831810-A2.
PD 23-JUL-1998.
PR 20-JAN-1998; U00218.
PR 21-JAN-1997; US-786624.
PA (SCHE ) SCHERING CORP.
PI Capone M, Gorman DM, Hedrick JA, Huffine CF, Rossi DL,
PI Vicari A, Zlotnik A;
DR P-PSDB; W69598.
DR Pure or recombinant chemokine CKDLR201.1 - useful, e.g. for treating
PT inflammation and as antiviral agents
PS Claim 7; Page 66-68; 77pp; English.
CC The present sequence encodes mouse G-protein coupled receptor 69A08 #2
CC protein which comprises a plurality of epitopes. Host cells containing
CC vectors comprising a nucleotide sequence encoding the protein are used to
CC produce recombinant protein. Treatment of a cell, particularly neuron,
CC macrophage or lymphocyte, carrying a G-protein coupled receptor (GPCR)
CC or a receptor responsive to CKDLR201.1 chemokine protein, with an
CC (ant)agonist is used to control physiological development, e.g.
CC alteration of calcium ion influx, a chemoattractant response, morphology,
CC phosphoinositide lipid turnover or an antiviral response. Nucleotide
CC sequences the CKDLR201.1 protein, or GPCRs, are useful as primers or
CC probes, e.g. for detecting and isolating related sequences and for
CC expressing antigenic peptides. Antibodies (Ab) directed against the
CC CKDLR201.1 protein and GPCRs are used to detect or purify the proteins;
CC diagnostically (e.g. for developmental abnormalities); in screening for
CC potential drugs; to inhibit chemokine/receptor activation; (when coupled
CC to a toxin or radioisotope) for killing specific cells, and to raise
CC anti-idiotypic antibodies. CKDLR201.1 protein and GPCRs and compounds
CC which bind them can be used to treat inflammation, e.g. asthma; as
CC antiviral agents, and to treat abnormal cell proliferation, regeneration,
CC degeneration and atrophy. Therapeutic agents are administered orally, by
CC injection and rectally.
SQ Sequence 1080 BP; 177 A; 347 C; 307 G; 249 T;
```

```
Query Match 12.6%; Score 618.2; DB 1; Length 1080;
Best Local Similarity 77.0%; Pred. No. 1.8e-72;
Matches 805; Conservative 0; Mismatches 233; Indels 7; Gaps 4;
```

```
QY 290 AGCGCCTCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 349
Db 42 AGTCTCTAGACAAGCTTAATCCAGAGGCTACCCGGGCAATCTGTGCCAACACAGT 101
QY 350 GACACCTCTGGAGCTCCCGACAGCTCAGGGGACCTGCTTCTGGGTGGGTGCCACACAGG 409
Db 102 GACAGCTGGAGCTCCCGGCGAGCTCTCAAGCACCTGCTGCTGGGTGGGTGCCACACAGT 159
QY 410 CTGGTGGCGCCCTCTATGGGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 469
Db 160 CTGGTGGTGGCTCTATGGGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 219
QY 470 CTGTGGGTGCTGGCCAGCAGGACCTGCTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 529
Db 220 CTGTGGGTGCTGGCCACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 279
QY 530 GCGACTGCTGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 589
```

```
Db 280 GCGAGTGGCTGATCTGCTGTTGGCCCTGGTGTGCCACACAGACTGGCTTACCACCTTCGCT 339
QY 590 GGGCAGGCTGGCCCTTCGGGAGGCGCCCTGGCCGCTGGCCAGCGCCGACACTCTATGCT 649
Db 340 GGGCAGGCTGGCCCTTCGGGAGGCTGGCTGGCCGCTGGCCAGCGCTGCCCCCTATGCT 399
QY 650 CACATGATGGCTCAGTGTGCTGCTGGCCGCTGACGCTGATGCTGATGCTGATGCTGCTG 709
Db 400 CACATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 459
QY 710 GTGACCCGCTGGGGCGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 769
Db 460 GTGATCTCTTTGGGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 519
QY 770 GCTGCTGGCTCAGTGGCGCCCTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 829
Db 520 GTGCGCTGGCTCTGTCGACGACCCCTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 579
QY 830 CGGCTGGCGGCTCCGATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 889
Db 580 C-GATTACTGGCTCCGATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 638
QY 890 TCCACTGCAACGGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 948
Db 639 TCCACTGGAAGCGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 698
QY 949 GGGCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1008
Db 699 GGGCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 758
QY 1009 CGGCGAGCGCTGAGGCTGACCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1068
Db 759 CAGCCATGCACTGAGCTGACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 818
QY 1069 CAGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1128
Db 819 TAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 878
QY 1129 CTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1188
Db 879 CTATGAGCCCTATGTGCCAGCTGGCAGCTGAGCAGCTGAGCAGCTGAGCAGCTGAGC 938
QY 1189 CATCTACTACTAGTGTGCGCCAGTTCAGGAGCAAGGTCGGGCGGCGAGGCTCTTCCAA 1248
Db 939 CATCTACTACTAGTGTGCGCCAGTTCAGGAGCAAGGTCGGGCGGCGAGGCTCTTCCAA 995
QY 1249 GTGCGCGGGGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1308
Db 996 CGAGCGGGAGGCGAGCAGTCTCTGAGGCTCCAGGAGGCTGGAAGCGGAGGAGTCTG 1055
QY 1309 CACCACTCTCTTGTGCTCCAGTGA 1333
Db 1056 CATTGTCTCTACACTCTCTGTGA 1080
```

```
RESULT 3
X37084/c
ID X37084 standard; DNA; 16891 BP.
AC X37084;
DT 06-JUL-1999 (first entry)
DE MEV gene sequence associated with familial mediterranean fever (FMF).
KW MEV; protein pyrin; Familial Mediterranean Fever; FMF; human;
OS FMF-associated mutant; hereditary disease; colchicine; ss.
PH Homo sapiens.
FT key Location/Qualifiers
FT CDS 1011..1450
FT /tag= a
FT /note= "contains introns"
FT exon 1011..1287
FT /tag= b
FT /number= 1
FT intron 1288..2806
```

```
FT /tag= c
FT /number= 1
FT 2807..3439
FT /tag= d
FT /number= 2
FT 3440..7812
FT /tag= e
FT /number= 2
FT 7813..8162
FT /tag= f
FT 8163..8591
FT /tag= g
FT /number= 3
FT 8592..8684
FT /tag= h
FT /number= 4
FT 8685..10346
FT /tag= i
FT /number= 4
FT 10347..10577
FT /tag= j
FT /number= 5
FT 10578..11044
FT /tag= k
FT /number= 5
FT 11045..11067
FT /tag= l
FT /number= 6
FT 11068..13002
FT /tag= m
FT /number= 6
FT 13003..13118
FT /tag= n
FT /number= 7
FT 13119..13304
FT /tag= o
FT /number= 7
FT 13305..13337
FT /tag= p
FT /number= 8
FT 13338..13700
FT /tag= q
FT /number= 8
FT /cons_splice= (5'site= Yes, 3' site=No)
FT 13701..13731
FT /tag= r
FT /number= 9
FT 13732..13896
FT /tag= s
FT /number= 9
FT 13897..14450
FT /tag= t
FT /number= 10
WO9909169-A1.
25-FEB-1999.
20-AUG-1998; U17255.
21-AUG-1997; US-056217.
(USSH ) US DEPT HEALTH & HUMAN SERVICES.
Aksentijevich I, Blake T, Centola M, Collins FS,
Deng Z, Doggett NA, Fischel-goodsian N, Gumucio DL,
Kastner DL, Liu PP, Pras M, Richards RL, Riecke DO,
Sood R;
WPI; 99-243555/20.
P-PSDB; Y09001.
A new gene, MEV, associated with Mediterranean Familial Fever
Claim 6; Fig 1; 49pp; English.
The present sequence represents a novel genomic nucleic acid sequence
(MEV) encoding the protein pyrin associated with Familial Mediterranean
Fever (FMF). Host cells transfected with a vector comprising a nucleic
acid encoding pyrin or its FMF-associated mutant, operably linked to a
functional promoter, are used for the recombinant expression of the
protein. The invention provides a method for diagnosing risk of FMF that
```

CC comprises analyzing a patient sample for an amino acid or nucleic acid
CC sequence of pyrin; and (b) correlating detection of a mutated sequence
CC with risk of developing FMF; Diagnostic kits developed in this invention
CC are used to identify and treat individuals at risk from FMF, a hereditary
CC disease prevalent in persons having a non-Ashkenazi Jewish, Armenian,
CC Arab, or Turkish background. Prior art FMF treatment with colchicine is
CC not effective in patients who are colchicine-resistant, and this
CC invention will cover all individuals.
SQ Sequence 16891 BP; 3905 A; 4308 C; 4316 G; 4356 T;

Query Match 9.4%; Score 458.6; DB 1; Length 16891;
Best Local Similarity 70.3%; Pred. No. 5.3e-52;
Matches 774; Conservative 0; Mismatches 279; Indels 48; Gaps 10;

QY 1402 CTCAGATGTGACCTATTGGAAAT-AGGGTTGTTACAACTGTCACT-----ACGGGAG 1455
DB 10067 CTCAGATGTGATCTATTGGAGATAGAGTTCCTGCAGATGTAATTAGTTAAGATGTG 10008
QY 1456 GTCACCTTTGGAGAAGGGTGGGCTTACATCCAGTGTGGTGGTGCCTCATAA--GATA 1512
DB 10007 GTCCTACTGGGATTGGTGGTCCCTAATCCCAATATAATGTGTCTATAAATGGGGA 9948
QY 1513 AGGAGAGGCCAGCCCTGTGGTCTACGCCCTGTAAATCCAGCACTTTAAGAGGCCAAGGCG 1572
DB 9947 AATTTGGGCTGGTGCATTTGGCTCATGTCTGTAAATCCAGCACTTTGGGAGGCTGAGTTG 9888
QY 1573 GATGATCACTTTAGCCAGGAGTTCAACACCCAGCCTGAGCAACATGTPAAACCCCATC 1632
DB 9887 GGCAGATCACTTGAGATCAGGAGTTGCGAGCCAGCCTGCCCCAATGCGGAACACTCTGTC 9828
QY 1633 TCTACCAAAATACAAAATTAAGCTGGGCTGTGGTGGTGGGCGCTGTATCCAGCTACT 1692
DB 9827 TTTACTAAAAATACAAAATTAAGCCAGGAGTGTGCGATCTGCCTACAGTCCCAGCACT 9768
QY 1693 CAGGAGACTGAGCAGAGGATCGCTTGAACCTGGGAGCAGAGG-TTGCAGTGAGCCGA 1751
DB 9767 CGGAGGCTGAGCGAGGAGAATTTGCTTCAACTGGGAGGCGAAGTTTGCAGTGAGCCGA 9708
QY 1752 GATTCGCCCACTGGACTCCAGCTCGCTGACAGAGAGCCTGCTCTAATTAATTAATTA 1811
DB 9707 GATTCAGCACTGCACCCAGCCTGAGCAGCGAGTGAGATTCAT------CTCAAAA 9654
QY 1812 ATTAATTAATTAATTTAAAAAGACGAAAAAGTGACGGCCAGGTGCGATGCTCACGCC 1871
DB 9653 AAAAAAATAAAAAAACCCTTAAAAAACAATAAAACCCAGCCAGCGATGCTCACACC 9594
QY 1872 TATATCTCAGCACTCTGGGAGGCC-AAGATGGAGGATTTGCTTGAAGCCAGAGTTGGG 1930
DB 9593 TGTATCCAGCACTTTGGGAGGCGGAGCAGGTGGATCACCTGAGGTGAGGTTTGAG 9534
QY 1931 ACCAGCCTGGGCAACATAGGGGATCCCATCTCTACACAAAAAATTTTAAATGAAC 1990
DB 9533 ACCAGCCTGGCCAAACAGGTGAATCCCTCTCTAC-----TAAATAACAAAAATTAGC 9479
QY 1991 CAGGCAATTTGGCATCGCCCTATAGTCCAGCCACTCAAGAGCCAGCGGAGGATCA 2050
DB 9478 CAGGCGCGTGGCGCGCTGTAATTCAGCCACTAAGAGGCTGAGCGCAGGAATTG 9419
QY 2051 CTTGAGCCTGGAGGTTGTGGTTCAGTGAGCTATGATTTGACACTGCATCCAGCCTG 2110
DB 9418 CTTGAACCTCGGAGGTGGAGTTGGAGCAAGCAAGATCGTGCCACTGTACTCCAGCCTG 9359
QY 2111 GCAACAGAGCAAGACTTTGTCT-----CAAAAAATAACAACCTAAAAATAAA 2159
DB 9358 GTGACAGGCGAGACTCCATCTTGGAAACAATAAACAACAGAGAGTAATTAATGAA 9299
QY 2160 AAAGAAGCAGAGATAGTGGTGTGGTCTCACACTGCAATCCAGCACTTTGGAG 2219
DB 9298 GTTAAAGTGTCTTTGGCGGCGCAGTGGCTCACATCTGTAATCCAGTACTTTGGGAG 9239
QY 2220 GCCGAGGTGGGAGATCATCTCAGGCCAGGAGTTCAAGACCAAGCCTGGCTACATGGTGA 2279

DB 9238 GCCAAGGGCGGGGATCA--CGAGTTCAGGAGTTCGAGACAGAGCTGACCAACATGGTGA 9181
QY 2280 AATCCTATCTCTACCAAAATACAAAATATTAGCCAGGCGTGGTGGCCACCTGTACTG 2339
DB 9180 AACCCCATCTCTACTAAACATACAAAATATTAGCCAGGCTGGTGGCACTACCTATAAT- 9121
QY 2340 GGGAGGTCCCAACCCAGCTACTCTGGGAGGCTGAGTTCAGGAGAAATCGCTTGAACCTGGAG 2399
DB 9121 -----CCAGCTACTCAGGAGCTGAGCAGGAGANTCGCTTGAACCCGGAG 9074
QY 2400 GGGAGGTTGGGTCAGCTGAGATGGTCCCACTGCTCCAGCCTGGCGGAAAGCCGAC 2459
DB 9073 GCAGAGTTGAGTGAGCGAGATCACACCACTGCTACACCTGGGTGACAGAGCGAG 9014
QY 2460 TCTGTCCTCAAAAAAAGAGA 2480
DB 9013 ACTCGCTCACAAAAAAGAAA 8993
RESULT 4
Q47355
ID Q47355 standard; cDNA; 1470 BP.
AC Q47355;
DT 09-FEB-1994 (first entry)
DE Myotonic dystrophy gene fragment containing deletion polymorphisms.
KW Myotonic dystrophy; disease; inherited; autosomal dominant; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT primer_bind 1..22
FT primer_bind /*tag= a
463..480
FT primer_bind /*tag= b
1449..1470
FT primer_bind /*tag= c
PN W09316196-A.
PD 19-AUG-1993.
PF 18-FEB-1993; CA0068.
PR 18-FEB-1992; US-837405.
PA (KORN/) KORNEJUK R G.
PA (MAHA/) MAHADEVAN M S.
PI Kornejuk RG, Mahadevan MS;
PI WPI; 93-272897/34.
PT Nucleotide molecule - comprises Myotonic Dystrophy locus of
PT chromosome 19q13, for diagnosis of disease status or risk
PS Disclosure; Figure 7; 58pp; English.
CC Myotonic Dystrophy is an inherited disease and is an autosomal
CC dominant disorder. It shows a marked variability in expression
CC ranging from a severe congenital form which is frequently fatal
CC to an asymptomatic condition associated with normal longevity.
CC An increase in the severity of the disease in successive generations
CC has been noted. The Myotonic Dystrophy gene (DM) encodes a protein
CC with putative serine-threonine protein kinase activity in normal
CC individuals. The increase in the severity of the disease in
CC successive generations is related to an increase in the number of
CC trinucleotide repeats in the DM gene. CTG repeats of up to 40
CC constitute a normal gene whereas repeats in excess of 40, especially
CC in excess of 50, constitute a mutant gene. The detection and
CC subsequent analysis of the number of the repeats region can be used
CC to provide information as to an individual's likelihood of developing
CC Myotonic Dystrophy. The sequence contains deletion polymorphisms.
SQ Sequence 1470 BP; 418 A; 325 C; 446 G; 281 T;

Query Match 9.2%; Score 448; DB 1; Length 1470;
Best Local Similarity 69.8%; Pred. No. 1.5e-50;
Matches 727; Conservative 0; Mismatches 285; Indels 50; Gaps 7;

QY 1514 GGAGAGGCCAGCCCTGGTGGCTCAGCCTGTATATCCCACTTTAAGAGGCCAAGCGG 1573
DB 153 GGGCTGGCGGGCGTGGTGGCTCAGCCCTGTATCTCAGCACTTTGGGAGCGCGGCGG 212
QY 1574 ATGATCACTTTGAGCCCAAGGAGTTCAACACCAAGCCTGAGCAACATGGTAAAAACCCATCT 1633

D	b	6145	GGTGGCGGCGCCTGTATGTCGCCAGCTACTTCGGGGAGGCTGAGGCAGAGAAATGGCTTGAAC	6086
Q	y	1724	CTGGGAGGCGAGGTTTGCACTGAGCCGAGATTGCGCCACTTGGACTTCAGAGCTGCGTGACA	1783
D	b	6085	CTGGGGCGGAGCTTGCAGTCAGCCGAGATCCCCTGCGTCCACTCTAGCCTGGGCGACA	6026
Q	y	1784	GAGAGCCTGTCTTAATTAATTAAATTAATTAATTAATTAATTAATTAATAAAGACAAAA	1843
D	b	6025	GAGCAAGACTCCAT-----CTCAAATAAATAAATAATTTAAAATAAATAAATACTTCCCT	5972
Q	y	1844	GTGACGGCCAGGTGCGAGTCGCGCTCACGCTATAATCTCAGCACTCTGGGAGGC--CAAGATG	1902
D	b	5971	G---AGGCCAGGTGGTGGCTCACACCTATAATCTCAGCACTTTGGAGGCTGAGGCGG	5915
Q	y	1903	GAGGATTGTTGAAGCCAGGAGTTTGGGACCAGCGCTGGGCAACATAGGGGATGCCCATCT	1962
D	b	5914	GTGGATCACTGAGTGCAGGAGTTCAAAACCAGCTGCGCCAACATGTTGAAACCTCGTCT	5855
Q	y	1963	CTACACACAAAAAATTTTTAATGAACAGGCATTGTGGCATCGGCATATAGTCCCAGC	2022
D	b	5854	CTAC-----CAAAAATACAAAAATTACCTGGGTGTGGGCACATGCTGTAACTCGGC	5800
Q	y	2023	CACCTAAGAGGCACAGCGGGAGGATCACTTGAGCCTGGGAGTGTGGTTGCAGTGAAGC	2082
D	b	5799	TACTTGGGAGGCTGAAGCAGGCAATCACTTGAGCCAGGAGCGGAGTTGCAGTGATC	5740
Q	y	2083	TATGATTACACTGCACTCCAGCCTGGGCAACAGAACAGACCTTGTCTCAAAATAA	2142
D	b	5739	CGAGATTACGGCCACCCCTACTCCAGCCTGGGTAAACAGAGTGAACATGTGTCTCAGAACAA	5680
Q	y	2143	ACAACTAAAAATTAAAAAAGAAGCAGAGATAGTGGGTGTGGTGCTCACACCTGCA	2202
D	b	5679	AAAA-----AAAGCCAGCGCTGGTGTCTCACGCTGTAA	5644
Q	y	2203	TCCAGCACTTTGGAAGCCGAGGTGGGCAGATCATCTGAGGCCAGGAGTTCAGACCCAG	2262
D	b	5643	TCCAGCACTTTGGGAGGCCAAGTGGGCGAATCA--CGAGGTTCAGGAGTTCAAGACCAG	5586
Q	y	2263	CTGGCTACATGGTCAATCTATCTCTACCABAAATACAAAATTAGCCAGCGTGCT	2322
D	b	5585	CTTGGCAATGTGTGAACCCCATCTCTACTAAATAACAAAATTAGTGGGTGGT	5526
Q	y	2323	GTGGGCACCTGTACTGGGAGGTGCCCAACCCAGCTACTGGGGAGGCTGAGTCAGAGAA	2382
D	b	5525	GGCATTCCTCTGAAT-----CCCAGTTACTTGGGAGGCTGAGGCAGGAA	5479
Q	y	2383	TGCCTTGAACCTGGGAGCGGAGGTTGGGGTCAGGTGNAGATGGTGCCACTGCACCTCCAG	2442
D	b	5478	TCACTTGAACCCGAGAGCGAGAGTTGCAGTGAGCCAGATTGCACCATTTGCACCTCCAGC	5419
Q	y	2443	CTGGGCGA--AAGAGCGCACTCTCTCTCCAAAAAAGAAGAGGAGGACACAGAGACA	2501
D	b	5418	CTGGGCAACAGGTGNAACTCTATCTCAGAAAAAANAANAATGTTCTCTGAGGAAA	5359
Q	y	2502	CACAGAG	2508
D	b	5358	CACAGAG	5352

RESULT 6

V35620/c

ID V35620 standard; DNA; 32367 BP.

AC V35620;

DT 07-SEP-1998 (first entry)

DE Human SHOX (short stature homeobox containing gene) gene sequence.

KW Homeobox domain; human growth gene; growth regulation; growth defect;

KW turner's syndrome; short stature homeobox containing gene; SHOXa;

KW SHOX; bone disease; osteoporosis; calcium regulation; short stature;

KW transcription factor A; ss.

OS Homo sapiens.
NC001450.1

PN WO9814568-A1.
00-2022-1000
00-2022-1000

PD 09-APR-1998.
DE 20-SEP-1007. 000000

PF 29-SEP-1997; E05355.

16-JAN-1997; EP-100583.
01-OCT-1996; US-027633.
(RAPP/) RAPPOLD-HOERBRAND G.
Rao E, Rappold-hoerbrand G;
WFI; 98-271719/24.
New human growth genes - used to develop products for the diagnosis
and treatment of human growth defects such as short stature, e.g.
Turner's syndrome
Claim 19; Pages 51-67; 84pp; English.
This is the human SHOX gene sequence containing the PAR1 region. The gene
region corresponding to short stature has been identified as a region of
approximately 500 kb in the PAR1 region of the X and Y chromosomes. Three
genes in this region have been identified as candidates for the short
stature gene. These genes were designated SHOX (also referred to as
SHOX33 or HOX93), pE92 and SHOT (SHOX-like homeobox gene on chromosome
three). The SHOX gene has two separate splicing sites resulting in two
variations SHOXA and SHOXB. The specification provides sequences of SHOX
(short stature homeobox-containing) genes SHOX EP92, SHOXA, SHOXB, SHOT
and exons of the SHOX genes as shown in V35610 to V35621 and protein
sequences of the human growth protein transcription factor SHOXA, SHOXB
and SHOT as shown W60573 to W60575. The novel genes are responsible for
human growth. Defects in the genes can cause short stature, e.g.
Turner's syndrome. The products can be used to develop agents for the
treatment of short stature or other human growth disorders. The products
can also be used for providing a mitogenic effect on cells, e.g. for the
treatment of bone diseases such as osteoporosis and diseases involved
with disturbance in the bone calcium regulation.
Sequence 32367 BP; 7627 A; 8130 C; 8564 G; 8043 T;
SQ

Query Match 8.2%; Score 402; DB 1; Length 32367;
Best Local Similarity 70.1%; Pred. No. 1e-44;
Matches 680: Conservative 0; Mismatches 230; Indels 60; Gaps 8;

[illegible]

2210	Qy	ACTTTGGGAAGCCGAGGTGGGCGAGATCATCTGAGGCCAGGAGTTCAGAACACGCGCTGGCT	2269
7016	Db	ACTTTGGGAGCGCGAGTGGGCGAGATCA--CGAGTTCAGGAGATCGAGACCATCTCGTGGCC	6959
2270	Qy	AACATGGTCGAATTCCTATCTCTACCAAAATACAAAAATTAGCCAGCGCTGGTGGTGGCG	2329
6958	Db	AACACAGTAAACCCCGTCTCTACTTAAAAATACAAAAATTAAGCTGGGCACGGTGGTGCAT	6999
2330	Qy	ACCTGTACTGGGAGGTGCCACCCAGCTACTTGGGGAGGCTGAGTCAGGAGAATCGCTTG	2389
6898	Db	GCCTGTA-----AACCCAGCTACCAGGTACTCGGGAGGCTGAGCAGGAGAATCGCTTG	6845
2390	Qy	AACCTGGGAGCGGAGTGTGGCTCAGCTGATGTTGCCACTGCATCCAGCTGGGCG	2449
6844	Db	AACCGAGGAGTCGGAGTTGCGGTGAGCTGAGATCATGCCACTGCATCGCGCTCGAGA	6785

Q7
2430 AAAAGGCGACCTGGTCTCAAAAAAAAAAAGAGGAGCAGA 2439
||||| | | | ||||| | | | | |
Db 6784 CAAGAGCAGACTCCGTCTCAAAAAAAAAAAGAGAAAAAA 6797

RESULT	10
ID	T78853/c
ID	T78853 standard; DNA; 6901 BP.
AC	T78853;
DT	06-FEB-1998 (first entry)
DE	Human lecithin-cholesterol acyltransferase (LCAT) DNA.
KW	Human lecithin-cholesterol acyltransferase; LCAT; lecithin; cholesterol
KW	Atherosclerosis; heart disease; stroke; heart attack;
KW	cholesterol reduction; peripheral vascular disease; Fish Eye Syndrome;
KW	myocardial infarction; Classic LCAT Deficiency Syndrome; ss.
OS	Homo sapiens.
FH	Key
FH	Location/Qualifiers
FT	1..812
FT	/tag= a
FT	/number= 1
FT	/note= "intron start not specified"
FT	misc_feature
FT	709..716
FT	/tag= b
FT	/note= "region indicated in figure 5, but no other information provided"
FT	misc_feature
FT	786..790
FT	/tag= c
FT	/note= "region indicated on figure 5, but no other information provided"
FT	misc_feature
FT	809
FT	/tag= d
FT	/note= "position is marked, but no further information is given"
FT	exon
FT	813..990
FT	/tag= e
FT	/number= 1
FT	837..5092
CDS	
FT	/tag= f
FT	/note= "includes introns"
FT	sig_peptide
FT	837..902
FT	/tag= g
FT	mat_peptide
FT	909..5029
FT	/tag= h
FT	/note= "introns included in sequence"
FT	intron
FT	991..1723
FT	/tag= i
FT	/number= 2
FT	exon
FT	1724..1880
FT	/tag= j
FT	/number= 2
FT	intron
FT	1881..1959
FT	/tag= k
FT	/number= 3
FT	exon
FT	1960..2075
FT	/tag= l
FT	/number= 3
FT	2076..2169
FT	intron

FT WO9814568-A1.
FT /tag= f
FT /note= "part of exon IV (G108)"
PD 09-APR-1998.
PN 29-SEP-1997; E05355.
PF 16-JAN-1997; EP-100583.
PR 01-OCT-1996; US-027633.
PA (RAPP/) RAPPOLD-HOERBRAND G.
PI Rao E, Rappold-Hoerbrand G;
DR WPI: 98-271719/24.
PT New human growth genes - used to develop products for the diagnosis
PT and treatment of human growth defects such as short stature, e.g.
PT Turner's syndrome
PS Disclosure; Pages 37-45; 84pp; English.
CC This is a preliminary nucleotide sequence of the SHOX gene. The gene
CC region corresponding to short stature has been identified as a region of
CC approximately 500 kb in the PAR1 region of the X and Y chromosomes. Three
CC genes in this region have been identified as candidates for the short
CC stature gene. These genes were designated SHOX (also referred to as
CC SHOX93 or HOX93), PTH92 and SHOT (SHOX-like homeobox gene on chromosome
CC three). The SHOX gene has two separate splicing sites resulting in two
CC variations SHOXa and SHOXb. The specification provides sequences of SHOX
CC (short stature homeobox-containing) genes SHOX E92, SHOXa, SHOXb, SHOT
CC and exons of the SHOX genes as shown in V35610 to V35621 and protein
CC sequences of the human growth protein transcription factor SHOXa, SHOXb
CC and SHOT as shown W60573 to W60575. The novel genes are responsible for
CC human growth. Defects in the genes can cause short stature, e.g.
CC Turner's syndrome. The products can be used to develop agents for the
CC treatment of short stature or other human growth disorders. The products
CC can also be used for providing a mitogenic effect on cells, e.g. for the
CC treatment of bone diseases such as osteoporosis and diseases involved
CC with disturbance in the bone calcium regulation.
SQ Sequence 15577 BP; 3692 A; 3875 C; 4196 G; 3629 T;

Query Match 7.1%; Score 345.8; DB 1; Length 15577;
Best Local Similarity 66.8%; Pred. No. 1.9e-37;
Matches 678; Conservative 0; Mismatches 212; Indels 125; Gaps 8;
QY 1518 AGCCAGGCTGTGCTCAGCCTGTATCCAGACACTTATAGAGGCCAAGCGGATG 1577
DB 10335 AGTCAGGTGGGTGCTCACCTCTATCCAGCACCTTTGGAGGCCAGGAGGTTG 10394
QY 1578 ATCATTTAGCCAGGAGTTCAACACAGCCTGAGCAATGTTAAACCCATCTCTAC 1637
DB 10395 ATCGCTTGGTCCAGGTTTGTAGACACAGCCTGGCCAAACATGTGAACCCGCTCTA- 10454
QY 1638 CAAAATACAAAATTTAGTGGCCTTGTGGCTGGCCCTGTATCCAGCTACTCAGGA 1697
DB 10454 CTAATAACGAAATTTAGCCAGGATGGTGTGAGCACCTGTATTCACGACTCTCGGA 10513
QY 1698 GACTGAGGAGAGGATCGCTTAACTGGAGGAGAGGTTGACGTGAGCGAGATTGC 1757
DB 10514 GGCTGAGGAGGAGATCGCTTGAACCTGGAGGTTGACCTGCTGAGTGAGTGATCAC 10573
QY 1758 GCACCTGGACTCCAGCCTCGGTGACAG-AGAGGCTGTCTCTAAATTAATTAAT 1813
DB 10574 ACCACTGCACTCCAGCCTGGATGAGAGAGCAAGCACTGTCTCAAAACAAATAAACA 10633
QY 1814 TANTTTTAAATTTTAAACAGCA-AGTACGCGGCGAGGTGCGAGTGG 1863
DB 10634 AAAAAACAAACAAATTTAAACAGCA-AGTACGCGGCGAGGTGCGAGTGG 10693
QY 1864 CTCACGCTTATATCTCAGCACCTCTGGAGG-CCAGATGGAGGATTTGTTGAAGCCAGG 1922
DB 10694 CTCACGCTTATATCTCAGCACCTTTGGAGGCCAGAGGTTGATGCTTGGAGTTCAGG 10753
QY 1923 AGTTTGGGACCGCTGGGCAACATAGGGGGATCCCATCTCTACACACAAATAATTTT 1982
DB 10754 AGTTGAGACCGCCTGGGCAACATAGTTGAACCCCATCTTTTACTAAATAACAAAGTT 10813
QY 1983 TAATGAACCGGATTTGTGGCATGGCTTATAGTCCAGCCACTCAAGAGGACAGCGGG 2042

DB 10814 -----AGCTGGGTGTGGTGTGGCCCTGTAAATCCAGCTACTCCGGAAGCTGAGGCTG 10868
QY 2043 GAGGATCACTTTCAGCCTGGGAGGTTGTGGTTCAGCTAGCTATGATTACCACTGCACT 2102
DB 10869 GAGAATTTGCTTGAATCTGGGAGGTGGAGTTGCAAGGAGGGGATAGTGCACATGCGAT 10928
QY 2103 CCAGCCTGGGCAACAGACAGCAAGACCTTGTCTCAAAATAAACAACATAAATAAATAA 2162
DB 10929 CCAGCCTGGGCAACAGACAGCAAGACCTTGTCTCAAAATAAACAACATAAATAA 10988
QY 2163 GAAGACGAGAGATAGTGGTGTGGTGTGCTACACCTGCAATCCAGCACTTTTGGAGGCC 2222
DB 10989 CAAG----- 10993
QY 2223 GAGGTGGGAGATCATCTCAGGCCAGGAGTTCAACAGCAGCCTGGCTACATGGTGAAT 2282
DB 10993 -----AGACCAGCCTGGCCACACATGGTGAAC 11019
QY 2283 CCTATCTCTACCAAAATACAAAATTTAGCCAGGCGTGTGGTGGGCACCTGTACTGGGG 2342
DB 11020 CGCGCTTTTACTAAAA--TACAAAATTAGCCGGGATGTGTGGGCACCTGTACT-- 11074
QY 2343 AGTGCCCAACAGCAGTCTGGGAGGCTGAGTCAGAGAAATCGCTTGAACCTGGAGGGG 2402
DB 11074 -----CCAGCTACTCTGGGAGGCTGAGGAGAGAAATGGCTTGAACCTGGAGGTG 11124
QY 2403 GAGGTGGGCTCAGCTGAGTGTGCTCCAGCTCCAGCCTGGCGGAGAAAGACACTCT 2462
DB 11125 GAGCTTGGAGTGGCGGAGATAGTCCACTGTCACCTGGCGGAGAGAGGAGGACT 11184
QY 2463 GTCTCCAAAAAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2517
DB 11185 TGATTTGAGACCCACCACCACCACCACCAACAAACAAACAAACAAACAAAC 11239

RESULT 13
T75284
ID T75284 standard; DNA; 5543 BP.
AC T75284;
DT 21-DEC-1998 (first entry)
DE Nucleotide sequence encoding human RAD50 intron 19, 3' end.
KW ds; human; RAD50; DNA repair; tumour suppression; cancer; Septin-2;
KW central nervous system.
OS Homo sapiens.
PN WO9727284-A2.
PD 31-JUL-1997.
PF 24-JAN-1997; U01299.
PR 17-JUL-1996; US-687080.
PR 26-JAN-1996; US-592126.
PA (GENE-) GENELABS TECHNOLOGIES INC.
PI Dolganov G.
PT WPI: 97-393672/36.
PT Human tumour suppressor gene RAD50 - useful to detect
PT predisposition to, decrease risk of and treat cancer, also Septin-2
PT homologues
PS Claim 1; Page 122-124; 195pp; English.
CC The human RAD50 (hRAD50) is involved in DNA repair and has tumour
CC suppression activity, can be used to detect predisposition to, decrease
CC the risk of or treat cancers, e.g. acute myeloid leukaemia,
CC myelodysplastic syndrome, therapy related myelodysplastic syndrome,
CC therapy related acute myeloid leukaemia, refractory anaemia or refractory
CC anaemia with excess blasts. Also disclosed in this invention is Human
CC Septin-2 homologues of which may be used as targets for cancer therapies
CC and central nervous system directed treatment methods, and to measure the
CC proliferative potential of selected cell types.
SQ Sequence 5543 BP; 1718 A; 1044 C; 1110 G; 1671 T;
Query Match 6.9%; Score 339.4; DB 1; Length 5543;
Best Local Similarity 73.8%; Pred. No. 1.4e-36;
Matches 492; Conservative 0; Mismatches 156; Indels 19; Gaps 4;
QY 1499 GTCTCATTAAGATAAGGAGAGGCCAGGCGCTGTGTGCTCACGCTGTATATCCAGCACTTT 1558

```
Db 630 GTAGTCAATTAAGAGGAGGAGCCAGGCGAGTCACTGCTGTATCCACCACTTT 689
Qy 1559 AAGAGGCGAAGCGGATGATCACTTGAGCCAGGAGTCAACACCACTGAGCAACAT 1618
Db 690 GGGAGGCGAAGCGGAGCGGATCACTGAGGTCAGGAGTTCAAGACTAGCTGGCCAAAT 749
Qy 1619 GGTAAACCCATCTCTACCAAAAATACAAAATAGCTGGGCTGGTGGCTGGCGCTG 1678
Db 750 GGTAAACCCATCTCTACTAGAAATAGAAAATAGCCAGGTGTGTGACACGCGCTG 809
Qy 1679 TAATCCAGCTACTCAGGAGACTGAGGAGAGGATGCTTGAACCTGGGAGGAGAGGT 1738
Db 810 TAATCCAGCTACTCAGGAGCTGAGGAGGAGGATGCTTGAACCTGGGAGGAGAGAT 869
Qy 1739 TGCAGTGGCGGAGATGCGCCACTGCGACTCCAGCCTGCGTGACAGAGAGCTGTCTTA 1798
Db 870 TGCAGTGGCGGAGATCAGACCATTCGACTTCAGCTGGGTGACA-AGAGTGAACCTCG 928
Qy 1799 AATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1858
Db 929 TCTCAAAACAAACAAACAA-----ACAAACAGAAACAAACAAACAGCGCGT 974
Qy 1859 AGTGGCTCACGCTATATATCTCAGCACTCTGGGAGGCGCAAGATGGAGTGTGTAAGC 1918
Db 975 GGTGGCTCACGCTGTATATCCAGCACTTTGGGAGGCGCAAGCGG-CCAGATCAGAGGT 1033
Qy 1919 CAGGAGTTGGGAGCGAGCTGGGCAACATAGGGGGATGCCATCTCTACACACAAAAAAT 1978
Db 1034 CAGGAGTCGAGACCATCTCGCTTAACACGCGCAACCCCGCTCTCTAC---TAAAAAATA 1090
Qy 1979 TTTTATGACACGAGCATTTGGCGATCGCGCTATATGCCAGGCTATGCCAGGCTCAGAGGCGACAG 2038
Db 1091 CAAAAAATTAACCAAGCGCTGTGTGGCGCTGTGTAGTCCAGCTACTCAGGAAGCTGAG 1150
Qy 2039 CGGGAGGAGTCACTTGAGCCTGGGAGGTTGTGTGAGTGTGAGTGTATGTATGACCTG 2098
Db 1151 GCAGGAGATGGCTGAACTTGGGAGGCGAGCTGTGAGTGTGAGGCGGAGATCGCACCTG 1210
Qy 2099 CACTCCAGCTGGGCAACAGAGCAAGACCTTGTCTCAAAAATTAACAACTAAATAA 2158
Db 1211 CACTCCAGCTGGGTGACAGAGCCAGACTCGTCTTAAAAAATAAAAAAAAAAAAAA 1270
Qy 2159 AAAGAA 2165
Db 1271 GTCACAA 1277
```

RESULT 14

```
Q75209 ID Q75209 standard; cDNA; 8342 BP.
AC Q75209;
DT 23-AUG-1995 (first entry)
DE ALI-1 (acute lymphocytic leukaemia-1) breakpoint cluster region.
KW Acute lymphoblastic leukaemia; acute nonlymphoblastic leukaemia;
KW chromosomal translocation; abnormality; detection; rearrangement;
KW breakpoint cluster region; Alu repeat; chromosome 11; probe B859; ds.
OS Homo sapiens.
FH Key
FT exon Location/Qualifiers
FT 1..263
FT /*tag= a
FT /number= 5
FT intron 264..2352
FT /*tag= b
FT 593..666
FT /*tag= c
FT /number= 6
FT intron 667..798
FT /*tag= d
FT 799..1108
FT /*tag= e
FT /rpt_type= OTHER
FT /note= "Alu repeat-a (Class J)"
```

```
FT repeat_unit 1119..1420
FT /*tag= f
FT /rpt_type= OTHER
FT /note= "Alu repeat-b (Class Sx)"
FT repeat_unit 1432..1716
FT /*tag= g
FT /rpt_type= OTHER
FT /note= "Alu repeat-c (Class Sb0)"
FT repeat_unit 1921..2216
FT /*tag= h
FT /rpt_type= OTHER
FT /note= "Alu repeat-d (Class J)"
FT exon 2353..2484
FT /*tag= i
FT /number= 7
FT intron 2485..3031
FT /*tag= j
FT exon 3032..3145
FT /*tag= k
FT /number= 8
FT intron 3146..6787
FT /*tag= l
FT repeat_unit 3973..4268
FT /*tag= m
FT /rpt_type= OTHER
FT /note= "Alu repeat-e (Class Sb0)"
FT repeat_unit 4764..5094
FT /*tag= n
FT /rpt_type= OTHER
FT /note= "Alu repeat-f (Class J)"
FT repeat_unit 6072..6362
FT /*tag= o
FT /rpt_type= OTHER
FT /note= "Alu repeat-g (Class S)"
FT exon 6788..6934
FT /*tag= p
FT /number= 9
FT intron 6935..7966
FT /note= "nucleotides 7429-7559 show about 80%
sequence identity to genomic sequences
found in 5' regulatory regions, 3' segments,
or in introns of several genes such as ApoA4,
Factor VIIIc subunit and G6PD"
FT repeat_unit 7164..7427
FT /*tag= q
FT /*tag= r
FT /rpt_type= OTHER
FT /note= "Alu repeat-h (Class Sx)"
FT exon 7967..8062
FT /*tag= s
FT /number= 10
FT intron 8063..8303
FT /*tag= t
FT exon 8304..8342
FT /*tag= u
FT /number= 11
FT W09426930-A.
FT 24-NOV-1994.
FT 22-APR-1994; U04496.
FT 14-MAY-1993; US-062443.
FT (UJJE-) UNIV JEFFERSON THOMAS.
FT Canaani E, Croce C;
FT DR P-PSDB; R66467.
FT New acute lymphocytic leukaemia gene prods. - used for the
FT diagnosis and treatment of leukaemias partic. acute
FT lymphoblastic or nonlymphoblastic leukaemia
FT Example 5; Fig 22; 207pp; English.
FT CC A phage clone, mg11.1, which spans the breakpoint cluster region in
FT CC the ALL-1 gene has been sequenced (Q75209). Eight Alu repeat
FT CC sequences were identified and classified based on criteria
FT CC published in Milosavljevic et al. (J.Mol.Evol. 32, 105-121, 1991).
FT CC The high concentration of Alu sequences within the area spanned by
```

CC exons 6 and 7 suggested a possible role for Alu in the chromosomal
CC translocations involving the ALL-1 gene. Homologous recombination
CC is not involved so the Alu repeats may act indirectly by
CC destabilising the region.

Query Match	6.9%	Score 336.4	DB 1	Length 8342
Best Local Similarity	67.5%	Pred. No. 3.2e-36		
Matches 666	Conservative 0	Mismatches 226	Indels 94	Gaps 10
Qy 1520	GCAGGCTGTGGCTCAGCGCTGAATCCCAAGCACTTTAAGAGGCCAGCGGATGGAT	1579		
Db 800	CTGGGACGGTGGCTCAGCGTGTATCCCAACACTTAGTGAGGCTGAGTGGGAGGAT	859		
Qy 1580	CACCTTGAGCCAGGAGTTCAACACAGCGCTGAGCAACATGTTAAACCCCACTCTCTACCA	1639		
Db 860	TGCTTGAGCCAGCAGTTCAAGACCAGCGCTGGCAACATAGCAAGACCCTGCTTTATTT	919		
Qy 1640	AAAAATACAAAA-----TTAGCTGGCTGTGGCTGGCGCTGTA	1680		
Db 920	AAACCAAAAAAAGAGAGAGAGAGATGTAGCCAGCATGTGGCATGTGGGTGTA	979		
Qy 1681	ATCCAGCTACTCAGGAGACTGAGCGAAGAGATCG-CTTGAACCTGGGAGGCAGAGTT	1739		
Db 980	GTCCAGCTACTCAGGAGGCTGAGATAGAAGATTGCTTGAGCCCAAGGAATCAAGGCT	1039		
Qy 1740	CGAGTGACCCGAGATTGCCCACTGGCACTCCAGCCTGGGTGACAGAGAGCCTGTCTTAA	1799		
Db 1040	GTAGTGAGCTATGATTGTACCCTCCAGCTCCAGCCTGGGTGACAAG-----	1087		
Qy 1800	ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1859		
Db 1087	-----CAAAACACTGTCTCCAAAAAAATTTAGGCTTGGCAAGCGCA	1129		
Qy 1860	GTGGCTCAGCGCTATAATCTCAGCACTCTGGGAGGCC--AAGATGAGGAGATGCTTGAAGC	1918		
Db 1130	CGCGCTCAGCGCTGTGATCCCAAGCACTTTGGGAAGCCCAAGCAGCAGATCACTTGAAGT	1189		
Qy 1919	CAGGAGTTTGGGACCAAGCTGGGCAACATAGGGGATCCATCTCTACACACAAAAAAT	1978		
Db 1190	CAGGAGTTGGAGACCAAGCTGGCAACATGGTGAACCTGTCTCTATGAAATACAA	1249		
Qy 1979	TTTTTAATGAACAGGAGATGTGGCATGGCCCTATAGTCCCAAGCAGCAGGACAG	2038		
Db 1250	AATTAGCCGGTGTGGTAGTGGTCTTGG-----TAATCTAGCTACTTGGGAGGCTGAG	1305		
Qy 2039	CGGGAGGA--TCACCTCAGGCTGGAGGTTGTGGTTCAGTGAGCTATGATTGTACCAC	2096		
Db 1306	CGAGGGGAGATTGCTGAAACCTGGAGCGGAGGCTGCAGTGAGCCAGATGTGATCAT	1365		
Qy 2097	TGCACCTCAGGCTGGGCAACAGCAGCAACCTGTCTCAAAATTAACAACTAAATTA	2156		
Db 1366	TGCACCTCAGCTGGCAACAGCAGCTAGCTCCATCCCAAAAAA-----	1420		
Qy 2157	AAAAAAGAGAGAGAGATAGTGGGTGTGGTGGCTCACACTGCAATCCAGCAGCTTTGG	2216		
Db 1420	-----AGTAGCGGCGCAGGTGGGTGAGCTGAGCTGATCCAGCAGCTTTGG	1464		
Qy 2217	AAGGCCGAGGTGGCGAGATCATCTCAGGCGCAGGAGTTCAAGACCAAGCCTGGCTAACATGG	2276		
Db 1465	GAGGCCGAGGCGGGGATCA--TGAGGCGAGGAGATCGAGACCATCTGGCTACACGG	1522		
Qy 2277	TGAATCCTATCTTACCACAAATAC-AAAAATTAGCAGCGGCTGGTGGGCACTGT	2335		
Db 1523	TGAACCTGTCTTACTAAAAATACAAAAATTAAGCCGGGAGGTGGCGGCGCTGT	1582		
Qy 2336	ACTGGGAGGTGCCCAACCCAGCTACTGGGAGGCTGAGTCAGGAGAAATCGCTTGAACCTG	2395		
Db 1583	AGT-----CCAGGCTACTCAGGAGGTGAGCCAGGAGAAATGGCTGAACCG	1629		
Qy 2396	GGAGCGGAGGTTGGGTCAGCTGAGATGGTGCCACTGCATCCAGCTGGCGGCAAGAG	2455		

Db 1630 GGSGGCGAGCCTGCAGTGAGCCGAGATCGCGCCACTGCACCTCCAGCTTGGGTGACACCG 1689

Qy 2456 CGACTCTGTCTCCAAAAAAGAGAGAA 2481
||||| ||||| ||||| |||

Db 1690 AGACTCCGCTCTCAAAAAAATAAA 1715

RESULT 15
V38933/C

ID	V38933	standard; DNA; 7146 BP.
AC	V38933;	
DT	24-SEP-1998	(first entry)
DE	Nucleic acid sequence of genomic DNA encoding human STCP-1.	
KW	Human; STCP-1; chemokine activity; T-cells; treatment; HIV infection;	
KW	inhibitory compound; assay; reduce; circulatory system STCP-1 level;	
KW	joint inflammation; rheumatoid arthritis; lupus; ds.	
OS	Homo sapiens.	
FH	Key	Location/Qualifiers
FT	CDS	1437..6204
FT		/*tag= a
FT		/product= STCP_1
FT		/note= "contains introns and exons"
FT	exon	1437..1509
FT		/*tag= b
FT		/number= 1
FT	exon	1510..3056
FT		/*tag= c
FT		/number= 1
FT	exon	3057..3180
FT		/*tag= d
FT		/number= 2
FT		3181..6119
FT	intron	/*tag= e
FT		/number= 2
FT	exon	6120..6201
FT		/*tag= f
FT		/number= 3
FN	W09824907-A1.	
PD	11-JUN-1998.	
PD	26-NOV-1997; U21552.	
PR	03-DEC-1996; US-760127.	
PA	(AMGE-) AMGEN INC.	
PI	Andrew DP, Chang M;	
DR	WPI; 98-333326/29.	
DR	P-PSDB; W62783.	
PT	Human STCP-1 polypeptides with chemokine activity - useful e.g. to	
PT	treat HIV infection or other viral or bacterial pathogens infecting	
PT	T-cells, macrophages or other immune system cells	
FS	Claim 1; Fig 2A-F; 96pp; English.	
CC	The present sequence encodes human STCP-1. STCP-1 polypeptides	
CC	demonstrate chemokine activity for T-cells. The polypeptides are useful	
CC	prophylactically or therapeutically to treat HIV infection and other	
CC	conditions associated with viral/bacterial pathogens infecting T-cells,	
CC	macrophages or other immune system cells. They can be included	
CC	(optionally chemically modified) with a pharmaceutically acceptable	
CC	carrier and optionally other pharmaceuticals (e.g. AZT, antibiotics etc.)	
CC	in therapeutic compositions for treating these conditions. STCP-1 also	
CC	useful to assay for inhibitory compounds used to reduce circulatory	
CC	system STCP-1 levels to alleviate e.g. joint inflammation associated	
CC	with rheumatoid arthritis, lupus or other autoimmune diseases. The	
CC	polypeptides are also useful to prepare antibodies or hybridomas. The	
CC	nucleic acids are useful to produce hybridisation probes to test for	
CC	STCP-1 DNA/RNA in mammalian samples.	
SQ	Sequence 7146 BP; 1664 A; 1814 C; 1883 G; 1785 T;	

	Query Match	6.88;	Score 331.2;	DB 1;	Length 7146;
	Best Local Similarity	72.08;	Pred. No. 1.5e-35;		
	Matches 508;	Conservative	0;	Mismatches 178;	Indels 20; Gaps 5;
Qy	1464	GGAGAAAGGTGGGGCTTACATCCAGTCTGGGTGTGTCTCTCATAGATPAAGGACGAGGCCA	1523		
Db	2607	GAAGGAGAGTTCAGGAATCCATGCTGAGACACAGGTTCTGGACAAAAATGGATAGTGAGGCC	2548		

QY 1524 GGCCTGGTGGCTCAGCCTGTAAATCCAGCAGCTTTAAGAGGCCAAGGGGATGGATCACT 1583
 Db 2547 AGAGCAGTGGCTCAGCCTGTAAATCCTAAACATTTAGGAGGCCAAGATGGGAGTTTCACT 2488
 QY 1584 TGAGCCAGGAGTTCAACACAGCAGCTGAGCAACATGTAAACCCCATCTCTACCAAAA 1643
 Db 2487 TGAGTCAAGAGTTGAGACAGACCCGTCAGCATGTGAACCCCATCTTTACTAAAA 2428
 QY 1644 TACAAAATATAGCTGGGCTTGGTGGCTGGGCGCTGTAAATCCAGCTACTCAGGAGACTGA 1703
 Db 2427 TACAAAATATAGCGGGCATGGTGGCAGTGCCTGTAAATCCAGCTACTCAGGAGGCTGA 2368
 QY 1704 GGCAGAGGATCGCTTGAACCTGGAGGCCAGAGGTTGCGAGTGAGCCGAGATTGCGCCACT 1763
 Db 2367 GGCAGGAGATCACTTTAACTCCAGAGGTAGAGTTGCGAGTGAGATTGCGCCACT 2308
 QY 1764 GGACTCCAGCCTCGGTGACAGAGAGCCTGTCTCTAAATTAATTAATTAATTAATTAAT 1823
 Db 2307 GCATCCAGCCTGGCACAGAGACTCAATCTCAAA--AATAAATAAATAAATAAATA 2250
 QY 1824 CAATTTAAAGACGAAAGTGAAGCGGCGAGTGCAGTGGCTCAGCCTATATCTCAGC 1883
 Db 2249 AAAAAATAA-----TTTAGGCTAGACGCGAGTGGCTCATGCTGTAAATCCAGC 2201
 QY 1884 ACTCTGGAGGCCAGAT--GGAGATTGCTTGAAGCCAGGAGTTGGAGCCAGCCTGGGC 1942
 Db 2200 ACTTTGGAGGCTGAGGTGGGAGGATCATCTGAGGTGAGGAGTTTGAGCCAGCCTGGCC 2141
 QY 1943 AACATAGGGGATCCCATCTCTACACACAAAAAATTTTTAAATGAACCGGCAATTTGG 2002
 Db 2140 AACATGTTGAACCTGTCTCCAC-----TAAAAATACAAAAATAGCAGGTGTGGTGG 2086
 QY 2003 CATGGCCTATAGTCCAGCCACTCAAGAGCCACAGCGGGGAGGATCACTTTAGCCTGGG 2062
 Db 2085 TGCATGCTGTAAATCCAGCTACTCAGGAGGCTCAGGAGGAGAAATCGCTTGAACCTGGG 2026
 QY 2063 AGTTTGTGGTTGCAGTGAGCTATGATTGTACCACTGCAGCTCCAGCCTGGGCAAC-AGAGC 2121
 Db 2025 AGGTAGAGGTTGCAGTGGGCCAGATTGCCCTCACTGCAGCTCCATCCTCTGGGCAACAGAGT 1966
 QY 2122 AAGACCTTGTCTCAAAAATAAACAACATAAATAAATAAATAAATAAATAAATAAATA 2167
 Db 1965 GAACTGTCTCAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1920

Search completed: October 29, 1999, 02:38:48
 Job time: 7875 sec

This Page Intentionally Left Blank

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 31, 1999, 00:45:31 ; Search time 1973.24 Seconds
(without alignments)
7889.367 Million cell updates/sec

Title: US-09-371-333-1
Perfect score: 4895
Sequence: 1 CCCCCACGGCTGGTGGCA.....AAAAAAAAAAAAAAAAAAAA 4895

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl.*

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_cm.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pr1.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_st.*
- 14: gb_sts.*
- 15: gb_sy.*
- 16: gb_un.*
- 17: gb_vi.*
- 18: em_fun.*
- 19: em_htg.*
- 20: em_hum1.*
- 21: em_hum2.*
- 22: em_in.*
- 23: em_om.*
- 24: em_or.*
- 25: em_ov.*
- 26: em_pat.*
- 27: em_ph.*
- 28: em_pl.*
- 29: em_ro.*
- 30: em_sts.*
- 31: em_sy.*
- 32: em_un.*
- 33: em_vi.*
- 34: gb_htg1.*
- 35: gb_htg2.*
- 36: gb_in1.*
- 37: gb_in2.*
- 38: em_ba1.*
- 39: em_ba2.*
- 40: em_hum3.*
- 41: em_hum4.*
- 42: gb_pr4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	4895	100.0	4895	11	AF055917	Homo sapi	

2	1531.4	31.3	1534	11	AF080214	Homo sapi
3	681.6	13.9	1361	12	AF080215	Mus muscu
4	503.2	10.3	173616	35	AC006509	Homo sapi
5	494	10.1	45328	10	AC002128	Human DNA
6	494	10.1	41936	10	CH19R30879	Homo sapi
7	489.2	10.0	145414	9	HS78F24	Homo sapi
8	489.2	10.0	134506	11	AC005738	Homo sapi
9	489	10.0	42416	11	AC005551	Homo sapi
10	486	9.9	122320	10	HS216E10	Human DNA
11	484	9.9	188477	34	AC004971	Homo sapi
12	484	9.9	109613	34	HS550H1	Homo sapi
13	482.8	9.9	199927	42	AC007055	Homo sapi
14	482.4	9.9	44957	11	AC005372	Homo sapi
15	481.2	9.8	44826	9	HS426N21	Human DNA
16	480	9.8	152083	35	AC005399	Homo sapi
17	480	9.8	148008	34	HS27Q5	Homo sapi
18	477.8	9.8	16689	10	HS4011712	Homo sapi
19	477.4	9.8	158905	11	AC005666	Homo sapi
20	477	9.7	104848	34	AC004588	Homo sapi
21	473.6	9.7	35587	11	AC005763	Homo sapi
22	473.6	9.7	45858	11	AC006123	Homo sapi
23	473.2	9.7	108523	42	AC000159	Homo sapi
24	473	9.7	40946	11	AC002984	Human DNA
25	472.8	9.7	109613	34	HS550H1	Homo sapi
26	471	9.6	133478	11	AC004686	Homo sapi
27	468.4	9.6	196439	34	AC005867	Homo sapi
28	467.2	9.5	80023	34	HS359F1	Homo sapi
29	467.2	9.5	169254	34	HS892F13	Homo sapi
30	467.2	9.5	169254	34	HS892F13	Homo sapi
31	466.2	9.5	126322	11	AC000026	Homo sapi
32	466.2	9.5	173029	42	AC002059	Homo sapi
33	465.8	9.5	46353	34	HS439A6	Homo sapi
34	465.6	9.5	91880	34	HS633019	Homo sapi
35	465.4	9.5	118899	10	HS1085F17	Human DNA
36	464.6	9.5	80141	10	HSW5GAR	Homo sapien
37	464	9.5	106921	42	AC004263	Homo sapi
38	463.8	9.5	57434	11	AC005808	Homo sapi
39	463.6	9.5	190508	42	AC005037	Homo sapi
40	463.4	9.5	39838	42	AC007766	Homo sapi
41	462.8	9.5	128328	11	AC005839	Homo sapi
42	462.6	9.5	128117	11	AC004000	Human PAC
43	462.2	9.4	162485	34	AC004840	Homo sapi
44	462.2	9.4	224895	34	HSC196A12	Homo sapi
45	461.6	9.4	152083	35	AC005399	Homo sapi

ALIGNMENTS

RESULT 1	AF055917	4895 bp	mrna	PRI	08-JUL-1998
LOCUS	AF055917	Homo sapiens protease-activated receptor 4 mRNA, complete cds.			
DEFINITION	AF055917	Homo sapiens protease-activated receptor 4 mRNA, complete cds.			
ACCESSION	AF055917	Homo sapiens protease-activated receptor 4 mRNA, complete cds.			
NID	3293321				
VERSION	AF055917.1	GI:3293321			
KEYWORDS	human.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	1	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	Xu, W., Andersen, H., Whitmore, T.E., Presnell, S.R., Yee, D.P., Ching, A., Gilbert, T., Davie, E.W. and Foster, D.C.				
TITLE	Cloning and characterization of human protease-activated receptor 4				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 95 (12), 6642-6646 (1998)				
MEDLINE	98283984				
REFERENCE	2	(Bases 1 to 4895)			
AUTHORS	Xu, W., Andersen, H., Whitmore, T.E., Gilbert, T., Davie, E.W. and Foster, D.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (26-MAR-1998) Biochemistry, University of Washington, J Wing, NE Pacific Street, Seattle, WA 98195, USA				

FEATURES
source

Location/Qualifiers
1. 4895
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19p12"
176..1333
CDS

/note="PAR4; member of family of G-protein coupled
protease activated receptors"

/codon_start=1

/product="protease-activated receptor 4"

/protein_id="AAC25699.1"

/db_xref="pid:g3293322"

/db_xref="gi:3293322"

/translation="MWGRLLWPLVLFSLSGTQTPSYDESSTGGDDTPSILP
APRGVGVCAANDSTLELPDSSRALLLWPTRLVPAALYGLVPLVGLPANGIALWVL
ATQAPRLSTMLLNATADLALALFPRIATHLRGQWFFGEAACKLATYALYGHM
YGSVLLAAVSLDRYLALVHLRRLALGLCMAAALPLTLQRTF
RLARSRLVCHDLALDQASHWOPAFCTLLALGCFLLPLAMLLCYKALPTLAASGR
RYGHALRTAVVLASAVAFVPSNLLLLHYSDPSPSAWNLGYAYVPSLALSLNSC
VDPEIYYVSAEFRDKVRAGLFRSPGDTVASKAEGSRGMGTHSSLQ"

BASE COUNT 971 a 1573 c 1400 g 951 t

ORIGIN

Query Match 100.0%; Score 4895; DB 11; Length 4895;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 4895; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CTCCACGGGCTGGTGGCAAGCGCCCTGGTGGGTCTCGGGGGGAGGGGAGCCTTCC	60
DB	1	CTCCACGGGCTGGTGGCAAGCGCCCTGGTGGGTCTCGGGGGGAGGGGAGCCTTCC	60
QY	61	TGGTTTATCTCACCGGCGGATCTGCTCGCTCGCTCGGCTCGAGAGCTGGGCTCAG	120
DB	61	TGGTTTATCTCACCGGCGGATCTGCTCGCTCGCTCGGCTCGAGAGCTGGGCTCAG	120
QY	121	GGTCCGGGAGGAGGAGCCTGAGGCCACAGCCAGCAGCAGCCTGAGTGCAGTGTG	180
DB	121	GGTCCGGGAGGAGGAGCCTGAGGCCACAGCCAGCAGCAGCCTGAGTGCAGTGTG	180
QY	181	GGGGGAGTCTTCTTGGCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	240
DB	181	GGGGGAGTCTTCTTGGCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	240
QY	241	CCCCAGCGCTTACGACGAGGAGGAGCAGCGAGGTGGTGTGATGACAGCACCCCTCAAT	300
DB	241	CCCCAGCGCTTACGACGAGGAGGAGCAGCGAGGTGGTGTGATGACAGCACCCCTCAAT	300
QY	301	CCTGCTGCCCCCGGGTACCCAGGCAAGTCTGTGCCAATGACAGTGACACCCCTGGA	360
DB	301	CCTGCTGCCCCCGGGTACCCAGGCAAGTCTGTGCCAATGACAGTGACACCCCTGGA	360
QY	361	GCTCCGGGAGCTCACGGGAGTCTTCTGGGCTGGTGGTGGTGGTGGTGGTGGTGGTGG	420
DB	361	GCTCCGGGAGCTCACGGGAGTCTTCTGGGCTGGTGGTGGTGGTGGTGGTGGTGGTGG	420
QY	421	CCTCTATGGGCTGGTCTGGTGGTGGGCTGCCGGCAATGGGCTGGGCTGGTGGTGGT	480
DB	421	CCTCTATGGGCTGGTCTGGTGGTGGGCTGCCGGCAATGGGCTGGGCTGGTGGTGGT	480
QY	481	GGCCAGGAGGAGCCTCGGCTGCCATCATGTGTGATGAACCTCGGAGCTGTGA	540
DB	481	GGCCAGGAGGAGCCTCGGCTGCCATCATGTGTGATGAACCTCGGAGCTGTGA	540
QY	541	CCTCTGCTGGGCTGGGCTGCCCGGGATCGCTACCACTGGCTGGGCGGCGCTG	600
DB	541	CCTCTGCTGGGCTGGGCTGCCCGGGATCGCTACCACTGGCTGGGCGGCGCTG	600
QY	601	GCCCTTCGGGAGGCGCTGCGGCTGGCCACGCGGAGCTCTATGTGTACATGTATGG	660
DB	601	GCCCTTCGGGAGGCGCTGCGGCTGGCCACGCGGAGCTCTATGTGTACATGTATGG	660

QY	661	CTCAGTGTCTGTGCTGGCGCCCTCAGCTGGATCGCTACCTGGCCCTGGCTGGTGCACCCGCT	720
DB	661	CTCAGTGTCTGTGCTGGCGCCCTCAGCTGGATCGCTACCTGGCCCTGGTGGTGCACCCGCT	720
QY	721	CGGGCCCGCCCTGCTGGCGCCCTGGCGCCCTGGACTCTGTCATGGCTGTGGCT	780
DB	721	CGGGCCCGCCCTGCTGGCGCCCTGGCGCCCTGGACTCTGTCATGGCTGTGGCT	780
QY	781	CATGGCGCCCGCCCTGGGACTGCCCCCTGACACTGAGGGGAGACCTTCGCGCTGGCGG	840
DB	781	CATGGCGCCCGCCCTGGGACTGCCCCCTGACACTGAGGGGAGACCTTCGCGCTGGCGG	840
QY	841	CTCCGATCCGCTGCTGCCATGACGCGCTGCCCTGACACACAGAGGCTCCCACTGGCA	900
DB	841	CTCCGATCCGCTGCTGCCATGACGCGCTGCCCTGACACACAGAGGCTCCCACTGGCA	900
QY	901	ACCGGCTTACCTGCTGGCGCTGTTGGGCTGTTCTGCCCCCTGCTGGGCACTGCTGCT	960
DB	901	ACCGGCTTACCTGCTGGCGCTGTTGGGCTGTTCTGCCCCCTGCTGGGCACTGCTGCT	960
QY	961	GTGCTACGGGGCCACCTGACACGCTGGCGCCAGCGCGCGCGGCTACGGCCACGGCT	1020
DB	961	GTGCTACGGGGCCACCTGACACGCTGGCGCCAGCGCGCGCGGCTACGGCCACGGCT	1020
QY	1021	GAGGCTGACCGCAGTGTGCTGGGCTCCGCGCTGGGCTTCTTCGTGGCCAGCAACCTGCT	1080
DB	1021	GAGGCTGACCGCAGTGTGCTGGGCTCCGCGCTGGGCTTCTTCGTGGCCAGCAACCTGCT	1080
QY	1081	GCTGCTGTGCTATCTGCGGACCGAGCCCGAGCGCTGGGGCAACCTCTATGTTGCCCTA	1140
DB	1081	GCTGCTGTGCTATCTGCGGACCGAGCCCGAGCGCTGGGGCAACCTCTATGTTGCCCTA	1140
QY	1141	CGTGCCAGCCTGCGCTGAGCACCTCAACAGCTGGTGGTACCTTCTATCTACTACTA	1200
DB	1141	CGTGCCAGCCTGCGCTGAGCACCTCAACAGCTGGTGGTACCTTCTATCTACTACTA	1200
QY	1201	CGTGTGCGCCGAGTTCAGGGACAAGTTCGGGCGAGGGCTCTTCCAACGGTGCCTGGGGGA	1260
DB	1201	CGTGTGCGCCGAGTTCAGGGACAAGTTCGGGCGAGGGCTCTTCCAACGGTGCCTGGGGGA	1260
QY	1261	CACGCTGCGCTCCAAAGGCTCTGCGGAAGGGGAGCGCGGGGATGGGCACTCCTC	1320
DB	1261	CACGCTGCGCTCCAAAGGCTCTGCGGAAGGGGAGCGCGGGGATGGGCACTCCTC	1320
QY	1321	TTTGCTCCAGTGACACAAGTGGGGAAGGCTGTACTGGTGGTGAACAGGGTCCCTTCCGCC	1380
DB	1321	TTTGCTCCAGTGACACAAGTGGGGAAGGCTGTACTGGTGGTGAACAGGGTCCCTTCCGCC	1380
QY	1381	ACTTCAGTCTCTTCTGGGACCTCAGAAATGACCTTATTTGGAATAGGGTTGTTACAA	1440
DB	1381	ACTTCAGTCTCTTCTGGGACCTCAGAAATGACCTTATTTGGAATAGGGTTGTTACAA	1440
QY	1441	CTGTCACTAGCGGAGGTACCTTTGGAGAGGGTGGGCTTACATCCAGTGTGGGTGGTGT	1500
DB	1441	CTGTCACTAGCGGAGGTACCTTTGGAGAGGGTGGGCTTACATCCAGTGTGGGTGGTGT	1500
QY	1501	CCTCATAGATAGGAGAGGAGGCGCTGGTGGTCTGACGCTGTAAATCCAGCACTTAA	1560
DB	1501	CCTCATAGATAGGAGAGGAGGCGCTGGTGGTCTGACGCTGTAAATCCAGCACTTAA	1560
QY	1561	GAGGCAAGGCGGATGATCACTTGGAGCCAGGAGTTCAACACAGCGCTGAGCAACATGG	1620
DB	1561	GAGGCAAGGCGGATGATCACTTGGAGCCAGGAGTTCAACACAGCGCTGAGCAACATGG	1620
QY	1621	TAAACCCCATCTCTACCAAAATACAAAATAGCTGGGCTTGGTGGCTGGCGCTGTA	1680
DB	1621	TAAACCCCATCTCTACCAAAATACAAAATAGCTGGGCTTGGTGGCTGGCGCTGTA	1680
QY	1681	ATCCAGCTACTCAGGAGCTGAGGCGAGAGGATCGGTTGAACCTGGGAGGAGGAGTTG	1740
DB	1681	ATCCAGCTACTCAGGAGCTGAGGCGAGAGGATCGGTTGAACCTGGGAGGAGGAGTTG	1740
QY	1741	CAGTGAGCCGAGATTGGCCACTGGACTCCAGCCTGCGTGCAGAGAGCCTGTCTCTAAA	1800

Qy	3961	AAGTGCAGGCGCAGCGGTCAGGACCGGGCCGGGCGGGGGTGGGGCGCGCGGGCCCTA	4020
Db	3961	AAGTGCAGGCGCAGCGGTCAGGACCGGGCCGGGCGGGGGTGGGGCGCGCGGGCCCTA	4020
Qy	4021	CCGGGTCGTAGTAGTGTACACGGAGACTGGCAGCGCCGAGCTCTGCCACACGAC	4080
Db	4021	CCGGGTCGTAGTAGTGTACACGGAGACTGGCAGCGCCGAGCTCTGCCACACGAC	4080
Qy	4081	TCCGGAGACAGCAACCGACGCTCAGGACCGGCTGGGATCTGTGGGCGACGG	4140
Db	4081	TCCGGAGACAGCAACCGACGCTCAGGACCGGCTGGGATCTGTGGGCGACGG	4140
Qy	4141	CGGGGCGAGCTCGACCGGGCCAGAGCGCGGGGCGCTGAGCTCAGGCGCCAGAACTG	4200
Db	4141	CGGGGCGAGCTCGACCGGGCCAGAGCGCGGGGCGCTGAGCTCAGGCGCCAGAACTG	4200
Qy	4201	GCTGATTCAGGATACCCAGGACGCGTGAACACAGAGAAACGTGATCCATTTCTT	4260
Db	4201	GCTGATTCAGGATACCCAGGACGCGTGAACACAGAGAAACGTGATCCATTTCTT	4260
Qy	4261	TTTTTCTTTTACTTTTCTTTTTTTTTTTTTTCTCAGACAGAGTCTCGCTGTGGCC	4320
Db	4261	TTTTTCTTTTACTTTTCTTTTTTTTTTTTTTCTCAGACAGAGTCTCGCTGTGGCC	4320
Qy	4321	AGGCTGAGTGCAGTGCAGTCTCGGCTCACTGCAAGCTCGGCTCTCGGTTCAAA	4380
Db	4321	AGGCTGAGTGCAGTGCAGTCTCGGCTCACTGCAAGCTCGGCTCTCGGTTCAAA	4380
Qy	4381	GATTCCTCTCGCTCAGCTCCCAAGTAGCTGGGATACAGGGCCCAACCGCACCTG	4440
Db	4381	GATTCCTCTCGCTCAGCTCCCAAGTAGCTGGGATACAGGGCCCAACCGCACCTG	4440
Qy	4441	CTAATTTTTTGTATTTTGTATCAAGACGAGTTTACCATTGTGGCGAGCTGTCTCA	4500
Db	4441	CTAATTTTTTGTATTTTGTATCAAGACGAGTTTACCATTGTGGCGAGCTGTCTCA	4500
Qy	4501	ACTCTCGCTCAAGTATCGGCTCGGTCCTATTTTATCTTTTGGGCTCTCCATCC	4560
Db	4501	ACTCTCGCTCAAGTATCGGCTCGGTCCTATTTTATCTTTTGGGCTCTCCATCC	4560
Qy	4561	CACGTGGAAAACGCTCTCAGTGCCTCTGAAACACACCTCTTTTGTGTGTGCAGCG	4620
Db	4561	CACGTGGAAAACGCTCTCAGTGCCTCTGAAACACACCTCTTTTGTGTGTGCAGCG	4620
Qy	4621	ATGCTCAGCATGTGTGGGTGGAGTCAGACATTCACGATATGTGCAATCATCACCTC	4680
Db	4621	ATGCTCAGCATGTGTGGGTGGAGTCAGACATTCACGATATGTGCAATCATCACCTC	4680
Qy	4681	TGCTAGTTACAGACGGTTCTTTCTCCCAAGAACCCGATCGCCATCAGCACTCA	4740
Db	4681	TGCTAGTTACAGACGGTTCTTTCTCCCAAGAACCCGATCGCCATCAGCACTCA	4740
Qy	4741	CTCCCACTCTCCCAAGCTCTGCAACCAAACTTTTCACTCTACGATTTGCCGT	4800
Db	4741	CTCCCACTCTCCCAAGCTCTGCAACCAAACTTTTCACTCTACGATTTGCCGT	4800
Qy	4801	TCTGGCATTTTCATGTCATGGAATCATGTACTCTGTGAAAAAATAAAAAAAAAA	4860
Db	4801	TCTGGCATTTTCATGTCATGGAATCATGTACTCTGTGAAAAAATAAAAAAAAAA	4860
Qy	4861	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	4895
Db	4861	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	4895
RESULT	2		
AF080214	AF080214	1534 bp	06-AUG-1998
LOCUS	Homo sapiens	protease-activated receptor 4 mRNA	PRI
DEFINITION	Homo sapiens	protease-activated receptor 4 mRNA, complete cds.	
ACCESSION	AF080214		
NID	93396080		
VERSION	AF080214.1	GI:3396080	

	complete sequence.	
ACCESSION	AC002128	repeat_region
NID	92121325	misc_feature
VERSION	AC002128.1 GI:2121325	
KEYWORDS	HTG.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 45328)	
AUTHORS	Lamerdin,J.E., McCreedy,P.M., Adamson,A.W., Burkhardt-Schultz,K., Garcla,E., Kyle,A., Ramirez,M., Stilwagen,S., Garnes,J., Danganan,L., Bruce,R., Quan,G., Montgomery,M., Ow,D., Kobayashi,A., Olsen,A.O. and Carrano,A.V.	
TITLE	Sequence analysis of a 1 Mb region of human 19ql3.1	
JOURNAL	Unpublished (1997)	
AUTHORS	2 (bases 1 to 45328) Lamerdin,J.E.	
TITLE	Direct Submission	
JOURNAL	Submitted (27-MAY-1997) Human Genome Center, Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94551, USA	
FEATURES	Location/Qualifiers	
source	1..45328	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="F19410"	
	/chromosome="19"	
	/map="19ql3.1 between D19S208 and CAPNS"	
	/maps="orientation is centromere to telomere"	
	/maps="overlaps with CH19R28588 to the left and CH19R30879 to the right"	
	/cell_line="UVSHL9-5B"	
	/cell_type="fibroblast"	
	/clone_lib="LL19NC02 F chromosome 19-specific cosmid library"	
	/note="cosmid library constructed at LLNL from flow-sorted chromosomes from hybrid UVSHL9-5B, which carries chromosome 19 as its only human chromosome"	
repeat_region	complement(146..433)	
repeat_region	/rpt_family="ALU"	
repeat_region	complement(1259..1543)	
repeat_region	/rpt_family="ALU"	
repeat_region	complement(2708..2987)	
repeat_region	/rpt_family="ALU"	
repeat_region	complement(3586..3816)	
repeat_region	/rpt_family="ALU"	
repeat_region	complement(3831..3904)	
repeat_region	/note="BLASTN similarity to T62858 (297..370); match: 0.85, score: 7.0e-144; database searched: est; yc03b08.s1 Homo sapiens cDNA clone 79575 3' similar to gb LI3712 HUMSCALUK Human scRNA molecule, transcribed from Alu"	
repeat_region	complement(3881..4165)	
repeat_region	/note="BLASTN similarity to T62858 (35..319); match: 0.99, score: 7.0e-144; database searched: est; yc03b08.s1 Homo sapiens cDNA clone 79575 3' similar to gb LI3712 HUMSCALUK Human scRNA molecule, transcribed from Alu"	
repeat_region	complement(3939..3989)	
repeat_region	/note="predicted exon, program: grail2exons_human_1.3, frame: 0, quality: good, score: 64.000"	
repeat_region	4068..4129	
repeat_region	/rpt_family="MLT1"	
repeat_region	complement(4152..4201)	
repeat_region	/note="BLASTN similarity to T62858 (1..50); match: 0.72, score: 7.0e-144; database searched: est; yc03b08.s1 Homo sapiens cDNA clone 79575 3' similar to gb LI3712 HUMSCALUK Human scRNA molecule, transcribed from Alu"	
repeat_region	complement(6334..6666)	
repeat_region	/note="predicted exon, program: grail2exons_human_1.3, frame: 1, quality: excellent, score: 89.000"	
repeat_region	6995..7293	
repeat_region	/rpt_family="ALU"	
repeat_region	7648..7891	
repeat_region	/note="BLASTN similarity to 264890 (1..185); match: 0.99, complement(29349..29533)	
repeat_region	/note="BLASTX similarity to (38..226); score: 6.6e-38; database searched: nr; hypothetical protein (L1H 5' region) - human"	
repeat_region	25337..26491	
repeat_region	/rpt_family="L1"	
repeat_region	26641..26930	
repeat_region	/rpt_family="ALU"	
repeat_region	27205..27450	
repeat_region	/rpt_family="L1"	
repeat_region	27460..27749	
repeat_region	/rpt_family="ALU"	
repeat_region	28059..28475	
repeat_region	/rpt_family="L1"	
repeat_region	complement(29349..29533)	
repeat_region	/note="BLASTN similarity to 264890 (1..185); match: 0.99, complement(29349..29533)	
repeat_region	/note="BLASTX similarity to (38..226); score: 6.6e-38; database searched: nr; hypothetical protein (L1H 5' region) - human"	
repeat_region	25337..26491	
repeat_region	/rpt_family="L1"	
repeat_region	26641..26930	
repeat_region	/rpt_family="ALU"	
repeat_region	27205..27450	
repeat_region	/rpt_family="L1"	
repeat_region	27460..27749	
repeat_region	/rpt_family="ALU"	
repeat_region	28059..28475	
repeat_region	/rpt_family="L1"	
repeat_region	complement(29349..29533)	
repeat_region	/note="BLASTN similarity to 264890 (1..185); match: 0.99, complement(29349..29533)	
repeat_region	/note="BLASTX similarity to (38..226); score: 6.6e-38; database searched: nr; hypothetical protein (L1H 5' region) - human"	
repeat_region	25337..26491	
repeat_region	/rpt_family="L1"	
repeat_region	26641..26930	
repeat_region	/rpt_family="ALU"	
repeat_region	27205..27450	
repeat_region	/rpt_family="L1"	
repeat_region	27460..27749	
repeat_region	/rpt_family="ALU"	
repeat_region	28059..28475	
repeat_region	/rpt_family="L1"	
repeat_region	complement(29349..29533)	
repeat_region	/note="BLASTN similarity to 264890 (1..185); match: 0.99, complement(29349..29533)	
repeat_region	/note="BLASTX similarity to (38..226); score: 6.6e-38; database searched: nr; hypothetical protein (L1H 5' region) - human"	
repeat_region	25337..26491	
repeat_region	/rpt_family="L1"	
repeat_region	26641..26930	
repeat_region	/rpt_family="ALU"	
repeat_region	27205..27450	
repeat_region	/rpt_family="L1"	
repeat_region	27460..27749	
repeat_region	/rpt_family="ALU"	
repeat_region	28059..28475	
repeat_region	/rpt_family="L1"	
repeat_region	complement(29349..29533)	
repeat_region	/note="BLASTN similarity to 264890 (1..185); match: 0.99, complement(29349..29533)	
repeat_region	/note="BLASTX similarity to (38..226); score: 6.6e-38; database searched: nr; hypothetical protein (L1H 5' region) - human"	
repeat_region	25337..26491	
repeat_region	/rpt_family="L1"	
repeat_region	26641..26930	
repeat_region	/rpt_family="ALU"	
repeat_region	27205..27450	
repeat_region	/rpt_family="L1"	
repeat_region	27460	

score: 3.0e-67; database searched: nt; H.sapiens CpG DNA, clone 176b10, forward read cp9176b10.ft1b . BLASTN similarity (1. .185); match: 1, score: 6.8e-70; database searched: Sanger CpG; bases 115 to 299 (SL to QR)" 30334. .>44147 /gene="Lisch7" /codon_start=1 join(30334. .30485,31780. .32114,40400. .40519,43380. .43436, 44001. .>44147) /gene="Lisch7" /note="Human homolog of mouse liver-specific gene, Lisch7.-BLASTN similarity to R12226 (1. .267); match: 0.91, score: 3.1e-54; database searched: est; vf52e09.r1.Homo sapiens cDNA clone 25701 5'.-BLASTN similarity to H42128 (48. .195); match: 1, score: 9.1e-54; database searched: est; y061b12.r1.Homo sapiens cDNA clone 182399 5'." /product="Lisch7" /protein_id="AAB58317.1" /db_xref="PIR:G2121326" /db_xref="GI:2121326" /translation="MQDGLGVGTRNGSGKRSVHPWPCAPRLRYFGRDARARA QTAAMALAIQVTSNPYHVVLPCTQYMTSTPTQPIVTKYKSFCDRIADA FSPASVNLQALNAAGNPGVNPVEQDSVRTVRVATKGNATVIGDYVQGRIT TGNADLFDQTAWGDGVSYYCSVYSQAQDLQGNNEAYAEILVLRGTSVAELLPFGQAG FIEDLFFVVCIAAFILFLLGLGICWCQCCPHTCCIVRCPCCPDCCCPA" complement(32251. .32519) /rpt_family="ALU" complement(33299. .33588) /rpt_family="ALU" 33831. .34121 /rpt_family="ALU" 34197. .34461 /rpt_family="ALU" 34504. .34789 /rpt_family="ALU" complement(35575. .35850) /rpt_family="ALU" 36519. .36778 /rpt_family="ALU" 36933. .37004 /rpt_family="ALU" 37416. .37704 /rpt_family="ALU" complement(38292. .38582) /rpt_family="ALU" 39633. .40077 /rpt_family="ALU" 40833. .41105 /rpt_family="ALU" complement(41747. .42045) /rpt_family="ALU" 42094. .42267

Query Match 10.1%; Score 494; DB 10; Length 45328;

Best Local Similarity 73.2%; Pred. No. 2e-72; Matches 748; Conservative 0; Mismatches 200; Indels 74; Gaps 6;

QY 1505 ATAAGATAAGAGAGCCAGCGCTGGTGCCTCAGCGCTGTAATCCAGCAGCTTTAAGAGG 1564
 DB 33808 ATGAGANTTCAAGTCTGGGACCGTGGCTCAGCTGTAATCTCAGCAGCTTTGGGAGG 33867
 QY 1565 CCAAGCGGATGGATCAGCTTAGCCCGCAGAGGTTCAACACCGCCTGAGCAACATGGTAA 1624
 DB 33868 CCAAGTGGTGGATCAGCTAGGTCAGGAGTTCAGACCGCCTGGCCCAACATGGTAA 33927
 QY 1625 ACCCATCTCTACCAAAAATACAAAATAGCTGGCTGGTGGCTGGCGCCCTGTAATCC 1684
 DB 33928 ACCCGCTCTCTACTAAAATACAAAATAGCCAGGATGGTGGCAGCATGCCGTGTAATCC 33987
 QY 1685 CAGCTACTCAGCAGACTGAGGAGAGAGGATCGCTTGAACCTGGGAGGAGAGGTTGCAGT 1744
 DB 33988 CAGCTACTCAGGAGGCTGAGGATCGGATCGCTTGAACCGAGGAGGAGGTTGCAGT 34047

QY 1745 GAGCGAGATTCGCCACTGCACTCCAGCTCCAGCTGCGTGACAGAGAGCGCTGTCTCTAATAA 1804
 DB 34048 GAGCTGAGATCGCCACCACTGTACTGCACTCTGGCGCAGAGAGTGAGACTGTGTCTCAAA 34107
 QY 1805 TTAATTAATTAATTA-----ATTCAATTTTA 1831
 DB 34108 AAAAAAAAAAAAAAAAAAGACTCCGTCAGGTATTAAGATGTCAGAGAGTACTAGTGTG 34167
 QY 1832 AAAAGAGAAAAGTACGGCCAGGTGCACTGCGCTCACGCCCTAT-AATCTCAGCACTCTGG 1890
 DB 34168 CAAAGAAAATACACCAAGCTGGGTGCATTCGCTCATGCTGTAAATTTACAGCACTTTGG 34227
 QY 1891 GAGGCCAA-GATGAGAGATTCCTTGAAGCCAGGAGTTTGGACCAAGCCTTGGCAACATAG 1949
 DB 34228 GAGGCCAAGGAGGAGGATGATCTTGAAGCTAGGAGTTTGAAGCAGCCTGGACACAAAA 34287
 QY 1950 GGGATGCCATCTCTACACACA---AAAAATTTTAAATGAACCAAGCAGCTTGTGGCATG 2006
 DB 34288 TGAGACCCCATGCTACAAAAATTTTAAAAATTTAAAAATTTAGCTGGGCATGTGGCATG 34347
 QY 2007 CGCCTATAGTCCCAAGCAGCACTCAAGAGGCACAGCGGGAGGATCACTTTGAGCCTGGAGGT 2066
 DB 34348 TGCCTGTGCTCCCGCTGCTCAGGAGGCTGAGGTGGGAGGATTTGCTGGCTTGAGAGGT 34407
 QY 2067 TGTGTTGCACTGAGCTATGATTTGACCTGCACTCCAGCTGCGCTGGCAACAGAGCAAGAC 2126
 DB 34408 CAAGCCTTCAAGTATGATGCTGCGCCTGCACTTCCAGCCTGGGTGACAGAGTGAGAC 34467
 QY 2127 CTGTGCTCAAAAATAAACAACAACTAAATTAATAAAGAGAGAGAGATAGTGGGTGTGG 2186
 DB 34468 CTGTCT-----TGAATGAAAAGAAATAGGTGGCGGAG 34504
 QY 2187 TGGTTCACACCTGCAATCCAGCACTTTGGAAGCGCGAGGTGGGAGATCATCTGAGGCC 2246
 DB 34505 TGGTTCACACCTGTAATCCAGCACTTTGGGAGCGCGAGGTGGGTGATCACCTGAGGTC 34564
 QY 2247 AGGATTCAGACCAAGCCTGGCTACATGTTGAATCTCTCTACCAAAAATACAAAA 2306
 DB 34565 AGGATTCAGACCAAGCCTGGCCCAACATGTTGAATCTCTCTCTACTAAAAATACAAAA 34624
 QY 2307 ATTAGCAGGCGTGTGTGGGCACTCTACTGGGAGGTGCCCAACCGCAGCTACTGGGGA 2366
 DB 34625 TTTAGCGGGGCTGGTGGTGGCGCTGTAT-----CCAGCTACTCTGGGA 34671
 QY 2367 GGCTGAGTCAGGAGAAATCGTTGAACCTGGGAGCGGAGTTGCGGTGAGTGGTGGATGGT 2426
 DB 34672 GGCTGAGCAGGAGAAATCGTTGAACCTGGGAGCGGAGTTGCGGTGAGTGGTGGATGGC 34731
 QY 2427 GCCACTGCACTCCAGCCTGGGCGAAGAGCGACTCTGTCTCAAAAAAGAGAGAGAGGA 2486
 DB 34732 GCCACTGCACTCTAGCCTGGGAAACAGTGAGACTCCGTCTTAAAAAAGAGAGAGAGGA 34791
 QY 2487 GA 2488
 DB 34792 AA 34793

RESULT 6

CH19R30879

LOCUS

DEFINITION

Accession

Version

Keywords

Source

Organism

Reference

Authors

Title

Direct Submission

CH19R30879 41936 bp DNA PRI 01-APR-1997
 Homo sapiens DNA from chromosome 19-cosmid R30879 containing USF2,
 genomic sequence.

AD000684
 GI:1905917

AD000684.1 GI:1905917
 chromosome 19; transcription factor.

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 41936)
 Lamerdin,J.E.

Direct Submission

JOURNAL Submitted (10-DEC-1996) J.E. Lamerdin, Human Genome Center,
Lawrence Livermore National Laboratory, 7000 East Ave, Livermore,
CA, USA, 94551 jane@cgst.llnl.gov ow@cornak.llnl.gov
COMMENT GSDB:S:1010600.
Human Genome Center
Biology and Biotechnology Research Program
Lawrence Livermore National Laboratory
7000 East Avenue
Livermore, CA 94550 USA
constructed at LLNL from flow-sorted chromosomes
from hybrid 5HL2-B, which carries chromosome 19 as its only human
chromosome.

FEATURES
source Location/Qualifiers
1. .41936
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/cell_line="5HL2-B"
/cell_type="fibroblast"
/clone="R30879"
/clone_lib="LL19NC03 R chromosome 19-specific cosmid
library"
complement(1. .34)
/note="repeat match = HSAL15306; putative"
/rpt_family="Alu"
repeat_region complement(811. .1099)
/note="repeat match = HSAL00208; putative"
/rpt_family="Alu"
repeat_region 1335. .1479
/note="repeat match = HSAL05512; putative"
/rpt_family="Alu"
repeat_region 1367. .2305
/note="repeat match = HSAL02043; putative"
/rpt_family="Alu"
repeat_region complement(3093. .3356)
/note="repeat match = HSAL02254; putative"
/rpt_family="Alu"
repeat_region 4030. .4168
/note="repeat match = HSAL03448; putative"
/rpt_family="Alu"
repeat_region 4031. .4288
/note="repeat match = HSAL01019; putative"
/rpt_family="Alu"
repeat_region 4435. .4533
/note="repeat match = HSAL02522; putative"
/rpt_family="Alu"
repeat_region 4449. .4565
/note="repeat match = HSAL01715; putative"
/rpt_family="Alu"
repeat_region 4926. .5218
/note="repeat match = HSAL06523; putative"
/rpt_family="Alu"
repeat_region complement(5804. .6091)
/note="repeat match = HSAL04878; putative"
/rpt_family="Alu"
repeat_region complement(6051. .6102)
/note="repeat match = HSAL04290; putative"
/rpt_family="Alu"
repeat_region 7131. .7434
/note="repeat match = HSAL02525; putative"
/rpt_family="Alu"
repeat_region 7137. .7589
/note="repeat match = HSAL00829; putative"
/rpt_family="Alu"
gene 7911. .16824
/gene="LISCH7"
CDS join<7911. .8030,10891. .10947,11512. .11658,15325. .15498,
15802. .15933,16075. .16700,16790. .16824)
/gene="LISCH7"
/function="human homolog of Mus musculus B6CBA Lisch7
mRNA"
/note="putative"
/codon_start=2

/product="liver-specific bHLH-zip transcription factor"
/protein_id="BAB51178.1"
/db_xref="PID:g1905918"
/db_xref="GI:1905918"
/translation="ADLTFDQTAMGDSGVYCSVVSAODLQGNNEAYAEILVLRGTS
VAELPGFQAGPIEGWVYVVCVLAFLIFLLGICWCQCPCCTCCYVRCPCCPDKC
CCPEARYAAGKAATGVPSTIYAPSTYAHLSPAKTPPPAMIPMGPAYNGYGGYPGDV
DRSSGVSRYIOAQODDSMRVLYYMEKELANFDPSPGPGSPGSRVGRGMSVTSLSH
EDDRSRSPRGPAITPIDEWGGHSPSPRGWQEPAPAREQAGGWRARRRRARSYDA
LDDLTPTSTAESGSPSTNGSRKAYMPSPRSKDDLYDQDDSDFFPSRDPHYDDF
RSREPPAPPPSHHRTDRDNGSRGDLDPYDGRLLLEAVKKGSEERRRPHKEE
EAYYPAPPYPYSETDSQASRERRLKDLALSRESLVV"
8303. .8386
/note="repeat match = HSAL05819; putative"
/rpt_family="Alu"
repeat_region 8325. .8609
/note="repeat match = ALU; putative"
/rpt_family="Alu"
repeat_region complement(9257. .9546)
/note="repeat match = HSAL05084; putative"
/rpt_family="Alu"
repeat_region complement(9513. .9555)
/note="repeat match = HSAL00054; putative"
/rpt_family="Alu"
repeat_region 9599. .9759
/note="repeat match = HSAL17230; putative"
/rpt_family="Alu"
repeat_region 9639. .9813
/note="repeat match = HSAL06130; putative"
/rpt_family="Alu"
repeat_region 10359. .10498
/note="repeat match = HSAL06554; putative"
/rpt_family="Alu"
misc_feature 11510. .11657
/gene="LISCH7"
/note="similarity: gb|H42128|H42128; y061b12.r1 Homo
sapiens cDNA clone 182399 5'; putative"
11957. .12072
/note="repeat match = HSAL10157; putative"
/rpt_family="Alu"
repeat_region 11981. .12253
/note="repeat match = ALU; putative"
/rpt_family="Alu"
repeat_region 12168. .12279
/note="repeat match = HSAL03934; putative"
/rpt_family="Alu"
repeat_region complement(12334. .12610)
/note="repeat match = HSAL00279; putative"
/rpt_family="Alu"
repeat_region complement(12391. .12982)
/note="repeat match = HSAL02441; putative"
/rpt_family="Alu"
repeat_region complement(12645. .13088)
/note="repeat match = HSAL02504; putative"
/rpt_family="Alu"
misc_feature 16521. .16586
/gene="LISCH7"
/note="similarity: gb|R39945|R39945; yf52e09.s1 Homo
sapiens cDNA clone 25701 3'; similar to . . .; putative"
16573. .16699
/gene="LISCH7"
/note="similarity: gb|R36881|R36881; yf52a08.s1 Homo
sapiens cDNA clone 25687 3'. . . .; putative"
16577. .16699
/gene="LISCH7"
/note="similarity: gb|R39945|R39945; yf52e09.s1 Homo
sapiens cDNA clone 25701 3'; similar to . . .; putative"
16594. .16650
/gene="LISCH7"
/note="similarity: gb|H06137|H06137; y177c10.s1 Homo
sapiens cDNA clone 44249 3'. . . .; putative"
16646. .16699
/gene="LISCH7"

85040 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr22>

78F24 is from the library RPC11 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pcypac2>.

FEATURES

```
source
1..145414
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="78F24"
/map="q12.1-12.3"
/clone_lib="RPC11"

repeat_region
1..835
/notes="L1 repeat: matches 3208..4060 of consensus"
1219..1514
/notes="AluX repeat: matches 1..302 of consensus"
1575..1770
/notes="L1 repeat: matches 448..4647 of consensus"
1776..2076
/notes="L1 repeat: matches 1..302 of consensus"
2100..2247
/notes="AluX repeat: matches 1..302 of consensus"
variation
2108..2110
/notes="L1 repeat: matches 4688..4838 of consensus"
/replace="aga"
complement(2351..2553)
/notes="AluX repeat: matches 299..1 of consensus"
2390..2392
/notes="clone 934P1; ata in this entry; substitution"
/replace="aca"
2936..3262
/notes="L1 repeat: matches 1..336 of consensus"
complement(3286..3588)
/notes="AluJ repeat: matches 302..2 of consensus"
3589..4123
/notes="L1 repeat: matches 328..892 of consensus"
4354..4569
/notes="L1 repeat: matches 453..670 of consensus"
4334..4439
/notes="L1 repeat: matches 453..557 of consensus"
4618..4981
/notes="L1 repeat: matches 3159..3546 of consensus"
4995..5292
/notes="AluS repeat: matches 1..299 of consensus"
5298..5596
/notes="AluX repeat: matches 3..300 of consensus"
5638..5844
/notes="L1 repeat: matches 3650..3858 of consensus"
5804..6123
/notes="L1 repeat: matches 35..312 of consensus"
6124..6431
/notes="AluX repeat: matches 1..301 of consensus"
6485..6628
/notes="L1 repeat: matches 364..518 of consensus"
complement(6676..6972)
/notes="AluS repeat: matches 300..1 of consensus"
6994..7127
/notes="L1 repeat: matches 55..186 of consensus"
complement(7129..7424)
/notes="AluS repeat: matches 303..2 of consensus"
7426..8399
/notes="L1 repeat: matches 178..1183 of consensus"
8400..8700
/notes="AluJ repeat: matches 1..300 of consensus"
```

```
repeat_region
complement(8701..8991)
/notes="AluJ repeat: matches 288..2 of consensus"
8993..9139
/notes="L1 repeat: matches 1168..1306 of consensus"
9268..9489
/notes="L1 repeat: matches 1545..1766 of consensus"
9491..9800
/notes="AluX repeat: matches 1..302 of consensus"
9808..10255
/notes="L1 repeat: matches 1757..2224 of consensus"
complement(10258..10393)
/notes="AluJ repeat: matches 301..190 of consensus"
10399..10622
/notes="L1 repeat: matches 72..294 of consensus"
complement(10623..10670)
/notes="AluJ repeat: matches 288..241 of consensus"
complement(10676..10855)
/notes="AluS repeat: matches 177..1 of consensus"
10858..11035
/notes="L1 repeat: matches 2217..2418 of consensus"
11034..11373
/notes="L1 repeat: matches 529..874 of consensus"
complement(11390..11691)
/notes="AluX repeat: matches 302..1 of consensus"
11696..11749
/notes="L1 repeat: matches 863..920 of consensus"
complement(11857..12037)
/notes="L1 repeat: matches 259..47 of consensus"
12094..12204
/notes="L1 repeat: matches 2..118 of consensus"
complement(12202..12239)
/notes="L1 repeat: matches 4967..4931 of consensus"
12240..12267
/notes="L1 repeat: matches 141..93 of consensus"
complement(12275..12323)
/notes="L1 repeat: matches 893..763 of consensus"
12519..12974
/notes="L1 repeat: matches 4931..5390 of consensus"
12832..13729
/notes="L1 repeat: matches 1..893 of consensus"
13707..13803
complement(14081..141115)
/notes="AluJ repeat: matches 302..1 of consensus"
14461..14631
/notes="L1 repeat: matches 73..248 of consensus"
14912..14914
/notes="L1 repeat: matches 13..159 of consensus"
/replace="ggg"
complement(15293..15594)
/notes="AluX repeat: matches 302..12 of consensus"
complement(15636..15708)
/notes="L1 repeat: matches 146..74 of consensus"
16169..16304
/notes="L1 repeat: matches 13..159 of consensus"
16486..16488
/notes="L1 repeat: matches 13..159 of consensus"
/replace="ggg"
16557..16855
/notes="AluJ repeat: matches 1..301 of consensus"
17057..17473
/notes="L1 repeat: matches 4756..5191 of consensus"
17483..17760
/notes="L1 repeat: matches 24..302 of consensus"
17773..17970
/notes="L1 repeat: matches 5184..5390 of consensus"
17830..18397
/notes="L1 repeat: matches 1..574 of consensus"
18167..18169
/notes="L1 repeat: matches 1..574 of consensus"
/replace="tat"
18398..18517
/notes="L1 repeat: matches 1..574 of consensus"
```

```
/note="AluJo repeat: matches 174. .293 of consensus"
18331. .18669
/note="MER42C repeat: matches 1392. .1536 of consensus"
20015. .20282
/note="AluJb repeat: matches 31. .302 of consensus"
complement(20911. .21027)
/note="MER45 repeat: matches 126. .1 of consensus"
21688. .21689
/note="clone 934P1: tg in this entry; deletion"
/replace="ttg"
22955. .23288
/note="AluSx repeat: matches 1. .302 of consensus"
24620. .24701
/note="MIR repeat: matches 63. .142 of consensus"
complement(27187. .27368)
/note="AluSx repeat: matches 182. .1 of consensus"
complement(27370. .27454)
/note="MIR repeat: matches 164. .76 of consensus"
31215. .31217
/note="clone 934P1: cgt in this entry; substitution"
/replace="cat"
complement(31868. .32006)
/note="MIR repeat: matches 208. .69 of consensus"
32054. .32056
/note="clone 934P1: aag in this entry; substitution"
/replace="agg"
complement(32170. .32476)
/note="AluJo repeat: matches 298. .1 of consensus"

Query Match 10.0%; Score 489.2; DB 9; Length 145414;
Best Local Similarity 73.7%; Pred. No. 1e-71;
Matches 716; Conservative 0; Mismatches 198; Indels 58; Gaps 5;

QY 1510 ATAAGGAGGCCAGGCGTGTGTGCTCAGCCGTGTAATCCAGCAGCTTTAAGAGGCCAAG 1569
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40100 AAAAAATGAGCCAGGCGTGATGCTCAAGCCTATATCTTAGCAGCTTTGAGAGCGAGT 40159

QY 1570 GCGATGGATCAGTTCAGGAGCCAGGAGTTTACACCCAGCCTGAGCAATGTTAAACCCC 1629
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40160 GCAGGCGATCAGTTGAGTGGTGGAGAGTTGAGAGTACCCGCGCAAGATGTTGAACCT 40219

QY 1630 ATCTCTACCAAAATACAAAATAGTGGCTGTGGTGGCTGGCGCTGTAAATCCAGCT 1689
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40220 GTCTCTACCAAAATACAAAATAGTGGCTGTGGTGGCTGTAAATCCAGCT 40279

QY 1690 ACTCAGGAGCTAGGAGGAGGATCGCTTGAACCTGGGAGGAGGAGTTGCAGTGAGCC 1749
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40280 GCTCAGGAGGCTAGGAGGAGGATCACTTGAACCTGGAAGCAGGAGTTGCAGTGAGCT 40339

QY 1750 GAGATTGGCCAGCTGAGCTCGCTGCTGACAGAGAGCTGTCTCTAAATTAAT 1809
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40340 GAGATTGGCTACTGCAACCCAGCTGGGCGACAGAGAGCTGTCTCTAA----- 40391

QY 1810 TAATTAATTAATCAATTTTAAAGACGAAAGTACGCGCCAGGTGCGTGCCTCAGC 1869
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40391 -----AAATATAAATAAATAAATAAATGAGAAGCTGGTGGTGGCTCATG 40441

QY 1870 CCTATAATCTCAGACCTCTGGAGGCCAAGATGGAGGATTCCTTGAAGCCAGGAGTTGG 1929
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40442 CCTGTAATCCAGCAGCTTTGGAGGCCAAGAGTGGAGAGCTCCTTTGAGGAGGAGTTCAA 40501

QY 1930 GACCAGCTGGGCAACATAGGGGATCCATCTCTACACACAAAAAATTTTAAATGAA 1989
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40502 GACCAGCTGGGCAACATAGCAGACCCCTGTCTCTACAAAAAATAA-----AATTAG 40555

QY 1990 CCAGGATTTGGGATCGGCTATAGTCCAGCCAGCTCAAGAGCAGGCGGAGGATC 2049
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40556 CCAGGTTGGTGGTGTACCTGTAGTTCTAGGTACTTCGGAGGCTGAGGCGAGGATG 40615

QY 2050 ACTTGGAGCTGGGAGGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2109
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40616 CCTTGGAGCA-GGAGGCTCAAGAGCTGCAGTGAGCTATGCTTACACCACTGCACTCCACAT 40674
```

```
QY 2110 GGCAACAGAGCAAGACCTTGTCTCTCAAAATAAACAACAACTAAATAAATAAAGACG 2169
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40575 GGGTGACAAAGCAAGACCCCTGTCT-----TTAAAAATAAAAAA 40714

QY 2170 AGAGATAGTGGTGTGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2229
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40715 AATAGGCGGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 40774

QY 2230 GCAGATCATCTGAGGCGGAGGTTCAAGACCGAGCTGGCTTAACATGTTGAAATCTATCT 2289
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40775 GCAGATCACCTGAGTTCAGGAGTTTCAGACCGAGCTGGCCAAACATGTTGAAACCCCTCT 40834

QY 2290 CTACCAAAATACAAAATTTAGCAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2349
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40835 CTACTAAAAACAAAAAATTTGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 40884

QY 2350 CACCCAGCTACTGGGAGGCTGAGTCAGGAGAAATCGTTTGAACCTGGGAGCGGAGTTG 2409
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40884 --CCAGCTACTTGGGAGGCTGAGGAGGAGAAATCACTTGAACCTGGGAGCGGAGTTG 40941

QY 2410 CGGTAGCTGAGATGTTGCCACTGCACCTCCAGCTGGCGGAAAGAGCGAGCTGTCTCCA 2469
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40942 CAGTGAGCTCGGATTTGTTCATTGCACTCCAGCTGGGCAACAGACGAGCTCCATCTC 41001

QY 2470 AAAAAAAGAGAA 2481
||| ||| ||| |||
Db 41002 AAAAATAAATAA 41013

RESULT 8
AC005738/c 134506 bp DNA PRI 20-OCT-1998
LOCUS Homo sapiens chromosome 5, BAC clone 7g12 (LBNL H126), complete
DEFINITION sequence.
ACCESSION AC005738 AC004619 L81405 L81406 L81678 L81679 L81680 L81860 L81861
L81862 AC001042 AC001043 AC001520
NID 93687213
VERSION AC005738.1 GI:3687213
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
Kadner,K., Miguel,T., Miller,C., Pollard,S., Pollard,M.,
Rojeski,H., Subramanian,S. and Martin,C.H.
Sequencing of human chromosome 5
Unpublished
2 (bases 1 to 134506)
Ricke,D.O.
Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
Unpublished
3 (bases 1 to 134506)
Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
Kadner,K., Miguel,T., Miller,C., Pollard,S., Pollard,M.,
Rojeski,H., Subramanian,S. and Martin,C.H.
Direct Submission
Submitted (01-OCT-1998) Human Genome Center, DOE Joint Genome
Institute, Lawrence Berkeley National Laboratory, MS 74-157,
Berkeley, CA 94720, U.S.A.
4 (bases 1 to 134506)
Ricke,D.O.
Direct Submission
Submitted (13-OCT-1998) Los Alamos National Laboratory, DOE Joint
Genome Institute, TA43, HRL-1, LS-3, MS M888, Los Alamos, NM 87345,
USA
5 (bases 1 to 134506)
Ricke,D.O.
Direct Submission
Submitted (20-OCT-1998) Los Alamos National Laboratory, DOE Joint
Genome Institute, TA43, HRL-1, LS-3, MS M888, Los Alamos, NM 87345,
```

```
US
COMMENT      Sequence submitted by:
DOE Joint Genome Institute
FEATURES
source      Location/Qualifiers
1. .134506
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="5p"
/chromosome="5"
join(1. .29755,78169. .82395,123646. .134506)
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5p"
/clone="BAC 7g12"
/note="LBNL H126"
complement(748. .891)
/note="GRAIL 2 excellent exon, frame 2"
2635. .2694
/note="GRAIL 2 excellent exon, frame 2"
/rpt_family="MLT1"
repeat_region
complement(2883. .2994)
/rpt_family="MLT1"
repeat_region
complement(3940. .4205)
/rpt_family="Alu"
repeat_region
complement(5036. .5148)
/rpt_family="MIR"
repeat_region
9194. .9650
/rpt_family="MER42"
repeat_region
9745. .10340
/rpt_family="MER42"
repeat_region
10346. .10624
/rpt_family="Alu"
repeat_region
10674. .10876
/rpt_family="MER42"
repeat_region
12173. .12472
/rpt_family="Alu"
repeat_region
14779. .14955
/rpt_family="L1"
repeat_region
15163. .15487
/rpt_family="Alu"
repeat_region
complement(19591. .19738)
/rpt_family="MIR"
repeat_region
20251. .20549
/rpt_family="Alu"
repeat_region
complement(21622. .21774)
/note="GRAIL 2 excellent exon, frame 1"
complement(22628. .22797)
repeat_region
/rpt_family="MIR"
repeat_region
complement(23374. .23671)
/rpt_family="Alu"
repeat_region
24507. .24641
/rpt_family="MER7"
source
28893. .60995
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5p"
/clone="P1 210h6"
/note="LBNL H52"
complement(34480. .34672)
/rpt_family="MLT1"
repeat_region
complement(34676. .34821)
/note="GRAIL 2 excellent exon, frame 0"
complement(34837. .35181)
/rpt_family="THEI"
repeat_region
38746. .38854
/note="GRAIL 2 excellent exon, frame 1"
39274. .39561
/rpt_family="Alu"
repeat_region
39856. .40341
/rpt_family="MER1"
repeat_region
complement(40375. .40451)
/rpt_family="MLT1"
repeat_region

44067. .44382
/rpt_family="Alu"
repeat_region
44592. .44900
/rpt_family="Alu"
repeat_region
45213. .45338
/rpt_family="L1"
repeat_region
45576. .45708
/rpt_family="Alu"
repeat_region
46954. .47235
/rpt_family="Alu"
repeat_region
50715. .51019
/rpt_family="Alu"
repeat_region
51522. .51600
/rpt_family="MIR"
repeat_region
complement(51960. .52100)
/rpt_family="MIR"
repeat_region
complement(52323. .52389)
/rpt_family="MER7"
repeat_region
complement(52403. .52692)
/rpt_family="Alu"
repeat_region
complement(52752. .53008)
/rpt_family="Alu"
repeat_region
complement(53050. .53338)
/rpt_family="Alu"
repeat_region
complement(53346. .53568)
/rpt_family="MER7"
repeat_region
complement(55100. .55354)
/rpt_family="L1"
repeat_region
complement(55478. .55533)
/rpt_family="Alu"
repeat_region
complement(56376. .56704)
/rpt_family="L1"
repeat_region
57025. .57175
/rpt_family="MLT1"
repeat_region
complement(58128. .58416)
/rpt_family="Alu"
repeat_region
complement(58603. .58884)
/rpt_family="L1"
repeat_region
59416. .59747
/rpt_family="MER21"
repeat_region
join(60081. .78733,82162. .124917)
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="P1 380h11"
/chromosome="5"
/map="5p"
/note="LBNL H122"
complement(60553. .60710)
/rpt_family="Alu"
repeat_region
complement(60718. .60914)
/rpt_family="THEI"
repeat_region
complement(61908. .62249)
/rpt_family="THEI"
repeat_region
complement(62279. .63781)
/rpt_family="MSTAR"
repeat_region
complement(63786. .64160)
/rpt_family="THEI"
repeat_region
65491. .65778
/rpt_family="Alu"
misc_feature
complement(66977. .67113)
/note="GRAIL 2 excellent exon, frame 2"
67119. .67157
/rpt_family="THEI"
repeat_region
complement(68013. .68120)
/note="GRAIL 2 excellent exon, frame 1"
68346. .68651
/rpt_family="MIR"
repeat_region
68894. .69226
/rpt_family="THEI"
repeat_region
complement(69684. .70023)
/rpt_family="THEI"
repeat_region
70567. .70681
/rpt_family="Alu"
repeat_region
```

```
repeat_region complement(72274..72537)
misc_feature /rpt_family="Alu"
repeat_region complement(74449..74802)
misc_feature /note="GRAIL 2 excellent exon, frame 2"
repeat_region complement(77047..77436)
misc_feature /rpt_family="MLR1"
misc_feature 79647..79829
misc_feature /note="GRAIL 2 excellent exon, frame 2"
misc_feature complement(82876..82937)
repeat_region /note="GRAIL 2 excellent exon, frame 0"
repeat_region 83530..83674
repeat_region /rpt_family="Alu"

Query Match 10.08; Score 489.2; DB 11; Length 134506;
Best Local Similarity 74.78; Pred. No. 1e-71;
Matches 728; Conservative 0; Mismatches 218; Indels 28; Gaps 8;

Qy 1518 AGGCGAGCCCTGGTGGCTCAGCCCTGTAATCCAGACCTTTAAAGAGGCCAAGCGGATGG 1577
Db 53347 AAGCCGGTGGTGGCTCAGCTCTGTAATCCAGACCTTTGGAGGCTGAGCGGGTGG 53288

Qy 1578 ATCACTTGAGCCCGAGGAGTTCAACACCCAGCCTGAGCAACATGGTAAACCCCATCTCTAC 1637
Db 53287 ATCACTTGAGGCTGAGGAGTTCTAGACCA-CCTGGCCACATGGTGAACCCCATCTGTAC 53229

Qy 1638 CAATAATCAAAAATAGTGGCTGGTGGCTGGCGCTGTAATCCAGCTACTCAGGA 1697
Db 53228 TGAATAATCAAAAATAGTGGCTGGTGGCTGGCTGTAATCCAGCTACTC--GG 53171

Qy 1698 GACTGAGCAGCAAGGATCGCTTGAACCTGGAGCAGAGGTTCAGTGAGCGGAGATTGC 1757
Db 53170 GCCTGAGCAGAGAAATGTTTGAACCGGAGGCGGAGGTTCAGTGAGCGGAGATTGC 53111

Qy 1758 GCCACTGACCTCCAGCCGCTGTCAGAGAGCCCTGTCTCTAAATTAATTAATTAAT 1817
Db 53110 ACCACTGACCTCCAGCCGCTGGCAACAGAGAGCTCCATCTCAAAAAAATAAAAAA 53051

Qy 1818 TTAATTCATTTTAAAGAGCAAAAGTGACGGCCAGCTGAGTGGCTCAGCGCTATAT 1877
Db 53050 AGAAATATCAGTAAGATACAGTATATATAGCCAGGCACTGTGGCTCATGACAGTAT 52991

Qy 1878 CTCAGCACTCGGAGGCAAGATGGAGGATGCTTGAAGCAGGAGTTTGGGACCGCC 1937
Db 52990 CACAGCACTTTGGAGGCAAGGAGGAG-CGATATCAGAGGTTCAGAGTTCAGACCGCC 52932

Qy 1938 TGGGCAACATAGGGGATGCCATCTCTACACACAAAAAATTTTATGTAACACCGCAT 1997
Db 52931 TGGCAACACAGTGAACCCCGCTCTCTCTAAAAATACAA---AAATCAGCCAGCGG 52876

Qy 1998 TGTGGCATGCCCTATAGTCCCGCACTCAAGAGGACAGCGGAGGATCACTTGAGC 2057
Db 52875 GGTGGCAGCGCCCTGTAGTCCCGCACTATGAGGAGCTGGAGGAGAGAGATGCTTGAAC 52816

Qy 2058 CTGGGAGTGTGGTTCAGTGAGTATGTTTACCACTGCACCTCCAGCTGGGCAACA 2117
Db 52815 CCAGGACAGAGGTTGAGTGGGCAAAATCTGCGCACTGCACCTCCAACTGGGCAACA 52756

Qy 2118 GAGCAAGACCTTGTCTCAAAAATTAACAACTAAATTAATAAAGAGAGCAGAGATAG 2177
Db 52755 GAGCAAGACTGTCTCAGAAAGAAA---GAAAGAAAAGAAAGATATAGTGGC 52700

Qy 2178 TGGGTGTGGTGCACACCTGCATCCAGCACTTTGGAGGCGGAGGTGGGAGATCA 2237
Db 52699 TGGGCGTGGTGGCTGACACCTGTAATCCAGCACTTTGGGAGGAGAGGTGGTGGATCA 52640

Qy 2238 TCTGAGGCGAGAGTTCAAGACCGCTGCTACATGCTGAATCTCTCTACCAAA 2297
Db 52639 --CAGGTGAGAGTTCAAGACCGCTGCTGAGGAGGAGGTGCTCTACTATAA 52582

Qy 2298 AATAC-AAAAATAGCCAGCGGTGGTGGGCGACCTTACTTGGGAGGAGTGCACCCAG 2356
Db 52581 AATACAAAAATAGCCAGCGGTGGTGGGCGGCTGTAGT-----CCAG 52535
```

```
Qy 2357 CTACTGGGAGGCTGAGTCAAGAGAACTCGTTGAACCTGGGAGCGGAGGCTGGGTGAG 2416
Db 52534 CTACTGGGAGGCTGAGGAGGAGAACTGGCGTGAACCTGGGAGCGGAGGCTGGCAGTGAG 52475

Qy 2417 CTGAGATGGTGGCCACTGCACCTCCAGCCTGGGCGAAGAGGAGCTCTCTCTCAAAAAA 2476
Db 52474 CCGAGATTTTGTCACTGCACCTCCAGCCTGGCAACAGAGGAGACTCCGCTCTCAAAAAA 52415

Qy 2477 GAGAAGAGGAGG 2490
Db 52414 AAAAAAGAAAAAG 52401

RESULT 9
AC005551 LOCUS 42416 bp DNA PRI 30-DEC-1998
DEFINITION Homo sapiens chromosome 19, cosmid R26529, complete sequence.
ACCESSION AC005551
NID G3482904
VERSION AC005551.1 GI:3482904
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 42416)
AUTHORS Lamerdin,J.E., McCready,P.M., Skowronski,E., Adamson,A.W.,
Burkhardt-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stilwagen,S.,
Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Danganan,L.,
Poundstone,P., Christensen,M., Georgescu,A., Avila,J., Liu,S.,
Attix,C., Andrise,T., Trankheim,M., Amico-Keller,G., Coefield,J.,
Auerle,S., Lucas,S., Bruce,R., Thomas,P., Quan,G., Krommiller,B.,
Arellano,A., Ow,D., Nolan,M., Trong,S., Kobayashi,A., Olsen,A.S.
and Carrano,A.V.
TITLE Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a
serine protease gene cluster
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 42416)
AUTHORS Lamerdin,J.E.
TITLE Direct Submission
JOURNAL National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT Map and sequence oriented from p telomere to centromere. Cosmid
R26529 is separated from cosmid F22676 to the left by approximately
1.5 kb, and is separated from cosmid R3374 to the right by
approximately 1.5 kb. Additional chromosome 19 map and sequence
information may be obtained at:
http://www-bio.llnl.gov/bbrp/genome/genome.html.
FEATURES
source
1..42416
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="R26529"
/chromosome="19"
/map="19p13.3 between CDC34 and D19S342"
/cell_line="5HL2-B"
/clone_lib="LL19NC03 R chromosome 19-specific cosmid
library"
/note="Cosmid library constructed at LLNL from flow-sorted
chromosomes from human-hamster hybrid 5HL2-B, which
carries chromosome 19 as its only human chromosome."
repeat_region 56..162
rpt_family="MIR"
repeat_region 187..473
rpt_family="AluX"
repeat_region 476..610
rpt_family="AluJo"
mRNA <834..11329
/gene="NFIC"
/product="H.sapiens mRNA for NF1/CAAT-binding
transcription factor 5 (CTF5)"
<834..11303
/gene="NFIC"
```

```

/product="Human mRNA for CCAAT-box binding transcription
factor CTF-1 (syn. CTF/NFI or CTF or NF-I or NF-1)"
<834. .11329
/gene="NFI"
/note="nuclear factor I/C (CCAAT-binding transcription
factor)"
join(<834. .1018,11104. .11121)
/gene="NFI"
/note="NFI /CAAT-binding transcription factor 5"
/codon_start=3
/product="CTF5"
/protein_id="AAC33192.1"
/db_xref="PID:G3482907"
/db_xref="GI:3482907"
/translation="TARSPHSSALHFTTSLPQTASTYFPHTAIRYPPHLPQDPL
KDLVSLADPASPQPGPSPKRWLG"
CDS
join(<834. .1018,2115. .2268,4902. .4987,11104. .11121)
/gene="NFI"
/note="CAAT BOX-BINDING TRANSCRIPTION FACTOR 1"
/codon_start=3
/product="NFI1_HUMAN"
/protein_id="AAC33191.1"
/db_xref="PID:G3482906"
/db_xref="GI:3482906"
/translation="TARSPHSSALHFTTSLPQTASTYFPHTAIRYPPHLPQDPL
KDLVSLADPASPQPGPSPKRWLG"
EGGATSPSPSPDTPANKRFVGLGRDPAGIYQAGSWLG"
CDS
join(<834. .1018,4902. .4952)
/gene="NFI"
/note="CAAT BOX-BINDING TRANSCRIPTION FACTORS 2 AND 3"
/codon_start=3
/product="NFI2_HUMAN"
/protein_id="AAC33190.1"
/db_xref="PID:G3482905"
/db_xref="GI:3482905"
/translation="TARSPHSSALHFTTSLPQTASTYFPHTAIRYPPHLPQDPL
KDLVSLADPASPQPGPSPKRWLG"
misc_feature
834. .1018
/gene="NFI"
/note="DDS similarity to N20996 yx57g02.s1 Homo sapiens
cDNA clone 265874 3' similar to gb:xl2492 CCAAT BOX-
BINDING TRANSCRIPTION FACTOR 1 (HUMAN); (482. .298); 96%
identity."
899. .1018
/gene="NFI"
/note="predicted exon, program: gail2exons_human_1.3,
frame: 1, quality: excellent, score: 89.000--(931. .1018)
DDS similarity to AA282279 zt12b08.r1 NCI_CGAP_GCB1 Homo
sapiens cDNA clone IMAGE:712887 5' similar to gb:xl2492
CCAAT BOX-BINDING TRANSCRIPTION FACTOR 1 (HUMAN); (1. .88);
100% identity.--Additional EST matches: AI083745"
1306. .1504
/rpt_family="AluJo"
repeat_region
1306. .1504
/rpt_family="AluJo"
repeat_region
1988. .2055
/rpt_family="GC_rich"
misc_feature
2115. .2268
/gene="NFI"
/note="DDS similarity to AA282279 zt12b08.r1 NCI_CGAP_GCB1
Homo sapiens cDNA clone IMAGE:712887 5' similar to
gb:xl2492 CCAAT BOX-BINDING TRANSCRIPTION FACTOR 1
(HUMAN); (89. .242); 100% identity.--Additional EST
matches: AI094574"
3292. .3389
/rpt_family="MIR"
repeat_region
3292. .3389
/rpt_family="MIR"
repeat_region
3753. .3785
/rpt_family="LINE2"
repeat_region
3786. .3795
/rpt_family="MIR"
repeat_region
4148. .4188
/rpt_family="MIR"
misc_feature
4902. .4987
/gene="NFI"
/note="DDS similarity to overlapping ESTs:-N20996

```

```

yx57g02.s1 Homo sapiens cDNA clone 265874 3' similar to
gb:xl2492 CCAAT BOX-BINDING TRANSCRIPTION FACTOR 1
(HUMAN); (297. .211); 97% identity.--AA282279 zt12b08.r1
NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:712887 5'
similar to gb:xl2492 CCAAT BOX-BINDING TRANSCRIPTION
FACTOR 1 (HUMAN); (243. .328); 100% identity."
complement(5259. .5352)
/rpt_family="MIR"
complement(5696. .5751)
/rpt_family="MIR"
misc_feature
6235. .6341
/gene="NFI"
/note="DDS similarity to AA282279 zt12b08.r1 NCI_CGAP_GCB1
Homo sapiens cDNA clone IMAGE:712887 5' similar to
gb:xl2492 CCAAT BOX-BINDING TRANSCRIPTION FACTOR 1
(HUMAN); (329. .435); 100% identity.--"
complement(6401. .6765)
/gene="NFI"
/note="DDS similarity to AA282217 zt12b08.s1 NCI_CGAP_GCB1
Homo sapiens cDNA clone IMAGE:712887 3'; Score: 704
Identity: 361/363 (99%).
complement(7093. .7146)
/rpt_family="LINE2"
8001. .8053
/rpt_family=" (GGGA)n"
repeat_region
8679. .8774
/rpt_family="MIR"
repeat_region
complement(8868. .9165)
/rpt_family="AluSx"
repeat_region
9195. .9493
/rpt_family="AluSx"
repeat_region
9524. .9790
/rpt_family="AluJo"
repeat_region
9793. .10094
/rpt_family="AluSx"
repeat_region
complement(10099. .10125)
/rpt_family="AT_rich"
repeat_region
10126. .10421
/rpt_family="AluSx"
repeat_region
10452. .10752
/rpt_family="AluSx"
repeat_region
10800. .10905
/rpt_family="MER53"
misc_feature
11104. .11314
/gene="NFI"
/note="DDS similarity to N20996 yx57g02.s1 Homo sapiens
cDNA clone 265874 3' similar to gb:xl2492 CCAAT BOX-
BINDING TRANSCRIPTION FACTOR 1 (HUMAN); (210. .1); 98%
identity.--Additional EST matches: AI094574, AI083745"
11286. .11334
/rpt_family=" (GAA)n"
repeat_region
complement(11504. .11534)
/rpt_family="GC_rich"
misc_feature
complement(11586. .12033)
/note="BLASTN similarity to AI089816 (15. .462); match:
0.98, score: 3.6e-180; database searched: month.na;
gailc09.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone
IMAGE:1686448 3' similar to contains MSRI.t3 MER22
repetitive.
12659. .13782
/note="DDS similarity to overlapping ESTs:-(12659. .13136)
AA497129 aa42c05.r1 Soares NHMPu S1 Homo sapiens cDNA
clone 823592 5'; Score: 751 Identity: 436/491
(88%).--(12899. .13296) N45485 yy59c07.r1 Homo sapiens cDNA
clone 277836 5'. Score: 706 Identity: 381/402
(94%).--(13506. .13074) AA071297 zm73d09.s1 Stratagene
neuroepithelium (#937231) Homo sapiens cDNA clone 531281
3'; Score: 709 Identity: 411/442 (92%).--(13782. .13308)
AA497062 aa42d05.s1 Soares NHMPu S1 Homo sapiens cDNA

```

Query Match 10.0%; Score 489; DB 11; Length 42416;
Best Local Similarity 74.2%; Pred. No. 1.3e-71;
Matches 722; Conservative 0; Mismatches 220; Indels 31; Gaps 7;

[illegible]

QY 1759 CCACTGGACTCCAGCTCGGTGACAGAGAGAGCTGTCTCTAAATTAATTAATTAATTAAT 1818
Db 87937 CCACTGGACTCCAGCTCGGTGACAGAGAGAGCTGTCTCTAAATTAATTAATTAATTAAT 1818
QY 1819 TAATTCATTTTAAAGACGACAAAGTGCAGCGGAGTGCAGTGCCTCAACGCTTAAATC 1878
Db 87898 -----CTGAAGAAAAAAGAGAGAGAGCTGTCTCTAAATTAATTAATTAATTAATC 1878
QY 1879 TCAGCACTCTGGAGGCGCAAGTGCAGTGCCTCAACGCTTAAATTAATTAATTAATTAATC 1938
Db 87845 CCAGCACTTTGGAGGCGGAGTGCAGTGCCTCAACGCTTAAATTAATTAATTAATTAATC 1938
QY 1939 GGGCAATAGAGGAGTGCAGTGCCTCAACGCTTAAATTAATTAATTAATTAATTAATC 1992
Db 87786 GGGCAATAGAGGAGTGCAGTGCCTCAACGCTTAAATTAATTAATTAATTAATTAATC 1992
QY 1993 GGCATGTTGGAGTGCAGTGCCTCAACGCTTAAATTAATTAATTAATTAATTAATTAATC 2052
Db 87726 GGCATGTTGGAGTGCAGTGCCTCAACGCTTAAATTAATTAATTAATTAATTAATTAATC 2052
QY 2053 TGAGCTGGAGGTTGTGTCTCAAGTGCAGTGCCTCAACGCTTAAATTAATTAATTAATTAATC 2112
Db 87666 TGAACCGGAGGTTGTGTCTCAAGTGCAGTGCCTCAACGCTTAAATTAATTAATTAATTAATC 2112
QY 2113 CAACAGCAAGAGCTTGTCTCAAGTGCAGTGCCTCAACGCTTAAATTAATTAATTAATTAATC 2172
Db 87606 TGACAGCAAGAGCTTGTCTCAAGTGCAGTGCCTCAACGCTTAAATTAATTAATTAATTAATC 2172
QY 2173 GATAGTGGTGTGTGTCTCAAGTGCAGTGCCTCAACGCTTAAATTAATTAATTAATTAATC 2232
Db 87547 ---ACCAGGCTGGTGTGTCTCAAGTGCAGTGCCTCAACGCTTAAATTAATTAATTAATTAATC 2232
QY 2233 GATCATCTGAGGCGGAGTGTGTCTCAAGTGCAGTGCCTCAACGCTTAAATTAATTAATTAATC 2292
Db 87490 GATCATCTGAGGCGGAGTGTGTCTCAAGTGCAGTGCCTCAACGCTTAAATTAATTAATTAATC 2292
QY 2293 CCAGCTACTGGGAGGCTGAGTGCAGGAGTGTGTCTCAAGTGCAGTGCCTCAACGCTTAAATC 2352
Db 87430 CTAAATATACAAATTAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2352
QY 2353 CCAGCTACTGGGAGGCTGAGTGCAGGAGTGTGTCTCAAGTGCAGTGCCTCAACGCTTAAATC 2412
Db 87383 CCAGCTACTGGGAGGCTGAGTGCAGGAGTGTGTCTCAAGTGCAGTGCCTCAACGCTTAAATC 2412
QY 2413 TCAGCTGAGT 2470
Db 87323 TGAGCGGAGATCGCAGT 2470
QY 2471 AAAAAAG 2516
Db 87263 AAAAAAG 2516
RESULT 11
AC004971/c AC004971 188477 bp DNA HTG 12-JUN-1998
LOCUS Homo sapiens clone DJ1125K23, WORKING DRAFT SEQUENCE, 21 unordered
DEFINITION pieces.
ACCESSION AC004971
NID 93213026
VERSION AC004971.1 GI:3213026
KEYWORDS HTG, HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 188477)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 188477)
AUTHORS Waterston,R.H.

TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order in this sequence record is
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1820: contig of 1820 bp in length
* 1821 1839: gap of unknown length
* 1840 3311: contig of 1472 bp in length
* 3312 3330: gap of unknown length
* 3331 5118: contig of 1788 bp in length
* 5119 5137: gap of unknown length
* 5138 6628: contig of 1491 bp in length
* 6629 6647: gap of unknown length
* 6648 8860: contig of 2213 bp in length
* 8861 8879: gap of unknown length
* 8880 10635: contig of 1756 bp in length
* 10636 10654: gap of unknown length
* 10655 13630: contig of 2976 bp in length
* 13631 13649: gap of unknown length
* 13650 17252: contig of 3603 bp in length
* 17253 17271: gap of unknown length
* 17272 20154: contig of 2883 bp in length
* 20155 20173: gap of unknown length
* 20174 24721: contig of 4548 bp in length
* 24722 24740: gap of unknown length
* 24741 30872: contig of 6132 bp in length
* 30873 30891: gap of unknown length
* 30892 36986: contig of 6094 bp in length
* 36987 37004: gap of unknown length
* 37005 40411: contig of 3407 bp in length
* 40412 40430: gap of unknown length
* 40431 50113: contig of 9684 bp in length
* 50115 50133: gap of unknown length
* 50134 67886: contig of 17753 bp in length
* 67887 67905: gap of unknown length
* 67906 84164: contig of 16259 bp in length
* 84165 84183: gap of unknown length
* 84184 105583: contig of 21400 bp in length
* 105584 105602: gap of unknown length
* 105603 128036: contig of 22434 bp in length
* 128037 128055: gap of unknown length
* 128056 185122: contig of 57067 bp in length
* 185123 185140: gap of unknown length
* 185141 186840: contig of 1700 bp in length
* 186841 186859: gap of unknown length
* 186860 188477: contig of 1619 bp in length.

FEATURES
Location/Qualifiers

source

1. 188477
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DJ1125K23"

BASE COUNT 43810 a 48919 c 48642 g 46728 t 378 others
ORIGIN

Query Match

Best Local Similarity 9.9%; Score 484; DB 34; Length 188477;
Matches 727; Conservative 0; Mismatches 245; Indels 36; Gaps 6;

QY 1507 AAGATAGGAG 1566
Db 94671 AAAATAAATGCTGGCTGGCGCCGCTGCTACACCTGTAATCCAGCAGCTTTAGAGGCC 94612
QY 1567 AAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1626
Db 94611 AAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 94552

```
QY 1627 CCATCTCTACCAAAATACAAAATAGCTGGCTGTGGCTGGCTGTAAATCCCA 1686
Db 94551 CCCATCTCTACTAAATAC-AAAATTAATCTGTGTGGCGCAGCCTGTAAATCCCA 94493
QY 1687 GCTACTCAGGAGCTGAGGCGAGAGTGCCTTTGAACCTGGGAGCAGAGTGTGCAGTGA 1746
Db 94492 GCTACTCGGAGGCTGAGGCGAGAGTGCCTTTGAACCCAGGAGTGTGTGCAGTGA 94433
QY 1747 GCGAGATGCGCCACTGGACTCCAGCTGCGTGCAGAGAGAGCCTGTCTCTAAATTAAT 1806
Db 94432 GCGAATACGCGCCACTGCTACCTCCAGCTGGGTGACAGATTGAGATC-----CATCT 94380
QY 1807 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1866
Db 94379 CAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 94320
QY 1867 AGCCCTATATCTCAGCACTCTCTGGAGGC-CAAGATGGAGGATTGCTTGAAGCCAGAGT 1925
Db 94319 AGCCCTGTAAATCCAGCACTTTGGAGGCTGAAGTATGTGATCACCCTGTGGCCAGAGT 94260
QY 1926 TTGGGACAGCCTGGCAACATAGGGGGATCCCATCTCTACACACAAAAATTTTAA 1985
Db 94259 TCAAGACCAAGCTGGCAACATAGGTGAACCCCTGTCTCTACTAAAAATAAA-----AAA 94205
QY 1986 TGAACAGGCTATGTGGCATGCGCCTATAGTCCAGCACTCAAGAGGCACAGCGGGAG 2045
Db 94204 AATCCGGGCTGTGGTGGGCACTTTATCCAGCTACTCGGAGCTGAAGCAAGAG 94145
QY 2046 GATCACTTGAGCCTGGAGGTTGTGTTGCAGTGAGCTATGATTGACCCTGCACCTCA 2105
Db 94144 AATTACTTGAACCTGGGAGCAGAGTGTGAGTAAGCCGAGATCTTGTCTGCCTCACTCA 94085
QY 2106 GCTGGGCAACAGCAAGCACTTGTCTCAAAATTAACAAATTAATTAATTAATTAAT 2165
Db 94084 GCTGGGCAACAGCAAGCACTTGTCTCAAAATTAATTAATTAATTAATTAATTAATTAAT 94025
QY 2166 GACGAGAGATAGTGGTGTGGTGGCTCACACCTGCAATCCAGCACTTTGGAGGCGGAG 2225
Db 94024 GGCCA-----GGCATGGTGGCTTACACCTGTAAATCCAGCACTTTGGAGGCGGAG 93974
QY 2226 GTGGGAGATCATCTGAGGCGAGGTTCAAGACACAGCTGGCTTAACATGGTGAATCTCT 2285
Db 93973 GCAGTGGATCATCTGAGTCAAGGTTCTAGACAGCCGCGCAATACGGTGAACCCCC 93914
QY 2286 ATCTCTACCAAAATACAAAATAGCCAGGCTGTGTGGCAGCTGTACTGGGGAGG 2345
Db 93913 ATTCTTACTAAATAACAAAATAGTGGACATGATGGCAGGTGCTCTAAT----- 93860
QY 2346 TGCCCAACCCAGCTACTGGGAGGCTGAGTCAGGAGAAATCGCTTGAACCTGGGAGCGGAG 2405
Db 93860 -----CCCATCTGCTCGGAGAGCTGAGGCATGAATCGCTTGAACCTTTGGGGTGGAG 93807
QY 2406 GTTGGGTCAGCTGAGATGGTGCACCTGCACCTCCAGCTGGGCGAAGAGCAGCTCTGTC 2465
Db 93806 GTTGAGTGAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 93747
QY 2466 TCCAAAAAAGAGAGGAGGAGGACACAGACACACAGAGAA 2513
Db 93746 TCTCAAAAAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 93699

RESULT 12
LOCUS HS50H1/c
DEFINITION Homo sapiens chromosome 20 clone 550H1, WORKING DRAFT SEQUENCE, in
unordered pieces.
ACCESSION AL035420
NID 95050978
VERSION AL035420.9 GI:5050978
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
```

```
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (10-JUN-1999) Wellcome Trust Genom Campus, Hinxtton,
JOURNAL Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
clonerequest@sanger.ac.uk
humqueres@sanger.ac.uk
COMMENT On Jun 11, 1999 this sequence version replaced g1:4741490.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Unfinished: dJ550H1 Contig_ID: 01651 acc=AL035420
Length: 74457 bp Unfinished: dJ550H1 Contig_ID: 01651 acc=AL035420
acc=AL035420 Length: 34356 bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
Location/Qualifiers
source
1..109613
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="550H1"
BASE COUNT 27126 a 25598 c 26628 g 29461 t 800 others
ORIGIN
Query Match 9.9%; Score 484; DB 34; Length 109613;
Best Local Similarity 74.5%; Pred. No. 7.6e-71;
Matches 724; Conservative 0; Mismatches 220; Indels 28; Gaps 8;
QY 1519 GCCCAGGCGTGTGGCTCAGCGCTGTAAATCCAGCACCTTTAAGAGCCCAAGCGCATGGA 1578
Db 95217 GCCCAGTGGTGGTACTCATCTGTCTACTCCAGCACCTTTGGAGCGCGGAGCGATGGA 95158
QY 1579 TCATTTGAGCCCGAGGAGTTCAACACCGCTGAGCAACATGGTAAACCCCATCTCTACC 1638
Db 95157 TCACCTGAGTTCAGGAGTTTGAACACCGCTGGCCAGCAGTGAATCTCATCTACT 95098
QY 1639 AAAATACAAAATAGCTGGCTGTGGTGGCGCTGTAAATCCAGCTACTCAGAG 1698
Db 95097 AAAATACAAAATAGCTGGCATGGTGGCACACGTTGTAGTCCAGCTACTTTGGGAG 95038
QY 1699 ACTGAGGAGAGAGTGCCTTGAACCTGGGAGGAGAGGTTGCAGTGAGCCGAGATTGCG 1758
Db 95037 GCTGAAGCAGGAGATCATTTGAACCTCAGGAGTTGGAGTTGCAGTGGCGCGAGATGGA 94978
QY 1759 CCATCTGAGCTCAGCGCTGCGTGACAGAGAGCGTGTCTCTAAATTAATTAATTAAT 1818
Db 94977 CCATCTGAGCTCAGCGCTGCGTGACAGAGAGCGCTGTCTCTCAAAACATAAAATAA 94918
QY 1819 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1878
Db 94917 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 94860
QY 1879 TCAGCAGCTCTGGAGGCGCAAGAT-GGAGGATGCTTTGAAGCCAGGAGTTTGGGAGCGCC 1937
Db 94859 CCAGCAGCTTTGGAGGCGCAAGTGGTGGTATCAGCTCAGGTCACAGTTCGAGACCGCC 94800
QY 1938 TGGGCAACATAGGGGATCCCATCTCTACACACAAAAATTTTAAATGAACAGGAT 1997
Db 94799 TGACCAACAGGGTAAACCCCGCTCTCTACTAAAAACAA-----AAAATAGCTGGCT 94744
QY 1998 TGTGCGCATCGCTATAGTCCAGCACTCAAGAGGCGACAGCGGAGGATCAGCTTGGC 2057
Db 94743 GGTGGCGAGTGCCTGTAAATCCAGCTACTTGGGAGGCTGAGGCGAGGAATCAGCTTGAAC 94684
QY 2058 CTGGGAGGTTGTGTGCGATGAGCTATGATTGTACCAGCTGCAGCTCCAGCTGGCAACA 2117
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 109613)
Skuce,C.

Direct Submission
Submitted (10-JUN-1999) Wellcome Trust Genom Campus, Hinxtton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
clonerequest@sanger.ac.uk
humqueres@sanger.ac.uk
On Jun 11, 1999 this sequence version replaced g1:4741490.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Unfinished: dJ550H1 Contig_ID: 01651 acc=AL035420
Length: 74457 bp Unfinished: dJ550H1 Contig_ID: 01651 acc=AL035420
acc=AL035420 Length: 34356 bp.

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

Location/Qualifiers
source

1..109613
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="550H1"

BASE COUNT 27126 a 25598 c 26628 g 29461 t 800 others
ORIGIN

Query Match 9.9%; Score 484; DB 34; Length 109613;
Best Local Similarity 74.5%; Pred. No. 7.6e-71;
Matches 724; Conservative 0; Mismatches 220; Indels 28; Gaps 8;

[illegible]

```

/clone_lib="RPCI-1"
1. .18094
/note="Overlap with BAC 316E14, found in AC006530, on the
3' side of 316E14. No sequence variations were found
between these two clones."
complement(join(<3486, 3558,5039, .5143,6745, .6830,
7509, .7607,11506, .11623))
/note="unknown; Intron-exon boundaries are defined in
relation to cDNA in L40399"
/codon_start=1
/protein_id="A4831936.1"
/db_xref="PID:g4885692"
/db_xref="GI:4885692"
/translation="MYVANKMTGLSTFIAPTEDIAQAACKDLTAVVADVLENSQIR
CQFPRDLVLPFLPRAARACTVMRQDNRTVDYTSVSSLSLASSEWNPVFARIEY
ADVSSQAESLAVKHNILYPYRTKGMHSMQVLQVDNKFIACLMSTKTBENGEA
"
complement(11506, .14848)
/note="This ORF is matched by some ESTs, N28386, AA6790504,
AA910059, AA128984, AA125983 and others. Blast x shows
similarity to PMS1 mismatch repair gene for a portion of
this ORF"
complement(18361, .29128)
/gene="ACVP1"
/note="acyl phosphatase"
complement(join(18573, .18788,28609, .28692))
/gene="ACVP1"
/note="acyl phosphatase; Intron-exon boundaries defined in
relation to cDNA in X84194, ESTs AA022534 and AI093586
give an additional exon at 26821-26899. There is a 5' UTR
from 29093-29128. Exon 2 begins at 28700"
/codon_start=1
/product="ACVP1"
/protein_id="A4D31937.1"
/db_xref="PID:g4885693"
/db_xref="GI:4885693"
/translation="MAEGNTLISVYEIFGVQGVFFPKHTQAEGKKLGLVGVWVQNTD
RGTVQGOGLQPIKSVRHMOWEITRGSPKSHDKANFNNEKVLKLDYSDFQIVK"
35751, .37067
/note="Open reading frame with no BLAST x similarity."
complement(join(49882, .49804,52136, .52348,53598, .53759,
56389, .56487))
/note="unknown; Intron-exon boundaries define in relation
to ESTs AA103140, AA182907, and AI555118"
/codon_start=1
/protein_id="A4D31938.1"
/db_xref="PID:g4885694"
/db_xref="GI:4885694"
/translation="M2ADRGMEGLISPTMGNNGASSCPGMLRKELENAEIPMP
DSPFLSAAFSSEKTELTYEELQGLVASEAPLSEHKPQVASEPRLNPVATVCAGKT
PLTPACASLSEQVERLQGLVLCALQAQKLOQENLIQTQLQKNLKKLEGSGQQVG
MHSKGTAKKEEMDPKDLDSQWCLGLTSDCRPSL"
complement(join(<60492, 60551,62216, 62386, 66111, .66272,
66776, .66884,68961, .69116,71089, .71115,71620, .71820,
72461, .72605,74787, .754979))
/note="unknown; Intron-exon boundaries defined by Genscan
and by ESTs AA8999979, AA568111, and D63222. The 5' and
3' ends of this gene could not be identified. The closest
BLAST x similarity is to the end of 'Her2.'"
/codon_start=2
/protein_id="A4D31939.1"
/db_xref="PID:g4885695"
/db_xref="GI:4885695"
/translation="SSTVTFVAVVTSFTSEVVYVWGSKTTPDKLDVIKSGCSAROV
CAGNTHFAVTVKEKLIYVWVWQGGTKLHGQCHGDKXASYRQPHKVEKQKLAHQVS
CGDDFTVCVTDGOLYAFSDYIGCMGVKDVAGPEVLSEPMFVLENPVQVSCGDN
HVVYLRNKVEYSGWGGEGYRGLDSIEEYTPQKVDYPAKIIIVAVQCGDGTFLLT
QSGVKLALNEFNKGLKQNMCSHDEHYTPVFTSFLAKQLSFYKTRITAPQK
THTAAIDERGLLTCFCNKCGOLGVGNYKKRLGILNGLGGLGGQVIRVSCGDDEFTTA
ADDNHILFANGNGNRRLAMVTPERPHGSDICTSPRIFGSLHHVPLDSCRWHITL
IVEKXHSKNTGSRNSGSLSTGT"
complement(join(<78441, .78505,79304, .79414,82298, .82429,
83933, .84038,85613, .85683,86222, .86277,89150, .89327,

```



```

REFERENCE      2 (bases 1 to 44957)
AUTHORS        Ricks,D.O.
TITLE          Large Scale Sequence Analysis and Annotation with the Sequence
                Comparison Analysis (SCAN) System
JOURNAL        Unpublished
REFERENCE      3 (bases 1 to 44957)
AUTHORS        Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
                Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
                Rojeski,H., Subramanian,S. and Martin,C.H.
TITLE          Direct Submission
JOURNAL        Submitted (01-AUG-1998) Human Genome Center, DOE Joint Genome
                Institute, Lawrence Berkeley National Laboratory, MS 74-157,
                Berkeley, CA 94720, U.S.A.
COMMENT        Sequence submitted by:
FEATURES       Location/Qualifiers
                source
                1..44957
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /map="5q"
                /clone="1195e2"
                /chromosome="5"
                /note="LBNL H73"
                212..680
                /rpt_family="Alu"
                /complement(1329..1587)
                /rpt_family="Alu"
                1625..1921
                /rpt_family="Alu"
                /complement(2061..2351)
                /rpt_family="Alu"
                2741..3016
                /rpt_family="Alu"
                /note="(T)24"
                3279..3302
                /rpt_type=tandem
                /rpt_unit=T
                /complement(3285..3903)
                /rpt_family="Alu"
                /complement(4010..4138)
                /rpt_family="Alu"
                4153..4176
                /note="(T)24"
                /rpt_type=tandem
                /rpt_unit=T
                /complement(4159..4437)
                /rpt_family="Alu"
                4483..4785
                /rpt_family="Alu"
                join(4854..4952,7649..7813,11526..11678,15096..15209,
                15884..16048,16908..17030,20748..20873,23801..23884,
                25757..25846,30119..30313,31286..31450)
                /standard_name="RNA helicase (Rat homolog)"
                /note="89-100% protein identity PIR:A57514"
                4856..4956
                /note="GRAIL 2 excellent exon, frame 2"
                /complement(5104..5336)
                /rpt_family="Alu"
                5373..5659
                /rpt_family="Alu"
                5913..6482
                /rpt_family="Alu"
                /complement(6552..6749)
                /rpt_family="Alu"
                /complement(6765..7042)
                /rpt_family="Alu"
                7649..7813
                /note="GRAIL 2 excellent exon, frame 1"
                /complement(8693..8801)
                /rpt_family="Alu"
                /complement(8811..9195)
                /rpt_family="Alu"
                /complement(9385..9732)

repeat_region  /rpt_family="Alu"
                complement(9842..10511)
                /rpt_family="Tigter2"
                10067..10357
                /rpt_family="Alu"
                complement(10526..10569)
                /rpt_family="Tigter2"
                10526..10569
                /rpt_family="MER8"
                10599..11178
                /rpt_family="Alu"
                11527..11678
                /note="GRAIL 2 excellent exon, frame 2"
                /complement(11974..12260)
                /rpt_family="Alu"
                complement(12535..12810)
                /rpt_family="Alu"
                complement(12860..13124)
                /rpt_family="Alu"
                complement(13152..13436)
                /rpt_family="Alu"
                complement(13489..13719)
                /rpt_family="Alu"
                13780..14047
                /rpt_family="Alu"
                14053..14098
                /note="(GT)23"
                /rpt_type=tandem
                /rpt_unit=GT
                14256..14500
                /rpt_family="Alu"
                14524..14563
                /note="(ATA)10"
                /rpt_type=tandem
                /rpt_unit=ATA
                14723..14999
                /rpt_family="Alu"
                15096..15297
                /note="GRAIL 2 excellent exon, frame 2"
                15884..16049
                /note="GRAIL 2 excellent exon, frame 1"
                16195..16467
                /rpt_family="Alu"
                complement(16478..16798)
                /rpt_family="Alu"
                16909..17029
                /note="GRAIL 2 excellent exon, frame 2"
                complement(17201..17504)
                /rpt_family="Alu"
                complement(17533..17790)
                /rpt_family="Alu"
                17829..18122
                /rpt_family="Alu"
                complement(18147..18462)
                /rpt_family="Alu"
                complement(18520..19276)
                /rpt_family="Alu"
                20757..20870
                /note="GRAIL 2 excellent exon, frame 2"
                complement(21133..21423)
                /rpt_family="Alu"
                21856..22123
                /rpt_family="Alu"
                22185..22208
                /note="(GT)8"
                /rpt_type=tandem
                /rpt_unit=GT
                complement(22254..22429)
                /rpt_family="Alu"
                23024..23330
                /rpt_family="Alu"
                23801..23879
                /note="GRAIL 2 excellent exon, frame 1"

```



```
repeat_region 35289..35587
/note="AluJb repeat: matches 1..296 of consensus"
repeat_region 35738..36047
/note="AluX repeat: matches 2..311 of consensus"
repeat_region 38259..38441
/note="MIR repeat: matches 8..192 of consensus"
repeat_region 38748..38774
/note="9 copies 3 mer gag 93% conserved"
misc_feature complement(<38992..>39426)
/note="match: GSS A0000811 clone 282A5"
repeat_region 41141..41332
/note="MIR repeat: matches 60..259 of consensus"
repeat_region 43144..43454
/note="AluX repeat: matches 2..312 of consensus"
repeat_region 43851..44046
/note="MIR repeat: matches 25..219 of consensus"
misc_feature complement(44597..44786)
/note="match: STS AL008917 clone 426N21"
BASE COUNT 10821 a 10356 c 11376 g 12273 t
ORIGIN

Query Match 9.8%; Score 481.4; DB 9; Length 44826;
Best Local Similarity 73.0%; Pred. No. 2.3e-70;
Matches 707; Conservative 0; Mismatches 241; Indels 21; Gaps 6;

QY 1519 GGCCAGGCGCTGGTGGCTCAGCGCTGTAATCCAGCACTTTAAGAGGCCAAGCGGATGGA 1578
DB 1519 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3587 GGCCGGCATGGTGGCCATGCTGTATCCAGCACTTTGGGAGGCTGAGGTGGGAAGGA 3528
DB 3587 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1579 TCACCTGAGCCAGGAGTTCACACCAAGCTGAGCAACATGGTAAACCCCATCTCTACC 1638
DB 1579 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3527 TCACCTGAGTCAAGAGTTCGAGACAGCGCTGGCCACATGGCCAAACCCCATCTCTACT 3468
DB 3527 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1639 AAAATACAAAATAGCTGGGCTGGTGGCTGGCGCTGTATCCAGTACTCAGGAG 1698
DB 1639 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3467 AAAATACAAAATAGCTGGGCTGGTGGCTGGCGCTGTATCCAGTACTCAGGAG 3408
DB 3467 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1699 ACTGAGGCAGAGGATCGCTTGAACCTGGGAGCAGAGTTCAGTGCAGCGGATTCGG 1758
DB 1699 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3407 GCTGAGGCAGAGATCGCTTGAAGCGGGAGGTGGAGGTTCAGTGCAGCGCAAGATAGCA 3348
DB 3407 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1759 CCAGTGCAGTCCAGCTGCTGACAGAGCCGTGCTCTAAATTAATTAATTAATTAAT 1818
DB 1759 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3347 CCAGTGCAGTTCAGCTGGGTGACAGAGCAAGACTCTGTCTAAACAAA----- 3297
DB 3347 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1819 TAATTCATTTTAAAGAGCAAGAGTTCAGCGCAGCTGCAGTGGCTCAGCCCTATAATC 1878
DB 1819 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3297 -AAAAAGAAACAGAGAGTTAAGTGAAGACCTGGCAGCGTGGCTCATGCTGTAAATC 3239
DB 3297 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1879 TCAGCACTCTGGGAGGCCAAGAT-GGAGGATTCCTTGAAGCCAGGAGTTTGGGACCGCC 1937
DB 1879 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3238 CCAACATGTTGGGAGGTGAGGTGGGAGATCACTTCAGTGCAGGAGTTCAGACCGCC 3179
DB 3238 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1938 TGGGCAACATAGGGGATCCCTCTCTACACAAAATAATTTTAAATGAACCCAGCAT 1997
DB 1938 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3178 TGGTCAACATGTTGAACCCCGTCTCTAC-----TAAATAACAGAGTTAGCCAGGCAT 3124
DB 3178 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1998 TGTGGCATGGCCCTATAGTCCAGCCACTCAAGAGCAGCAGCGGGAGGATCATTGAGC 2057
DB 1998 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3123 GGTGGTGGGCACGTTTAAATCCAGCTACTCAGGAGGCTGAGGTGGGAGATAGCTTGAAC 3064
DB 3123 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2058 CTGGGAGGTTGTTGGTGCAGTGCATGATTTAGCACTGCACCTCAGCTGGGCAACA 2117
DB 2058 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3063 CCAGGAGAGCGAGGTTGCGAGTGCAGGAGATCACACCACTGTACTCCAGCTGGGCAACA 3004
DB 3063 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2118 GAGCAGACCTTCTCAAAAATAAACAACACTAAATTAATAAAGAGACAGAGATAG 2177
DB 2118 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3003 GAGTAGACTGTCTCGGGTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2944
DB 3003 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2178 TGGGTGTTGGTCTCAGCTCAATCCAGCACTTTGGAAGCGCGAGGTGGGCAGATCA 2237
DB 2178 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2944 CGGGTGGGTGCTCAGCGCTGTATCCAGCACTTTGGGAGGCCAAGATGGGCAGATCA 2885
DB 2944 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 2238 TCTGAGGCCAGGATTCAAGACCAGCCTGGCTAACATGATGTAATCCTATCTCTACCAA 2297
DB 2238 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2884 --CAAGCTCAGGAGATTGAGACCCTCTGGCTTAACACGGTGAACCCCGTCTCTACTAAA 2827
QY 2298 AATACAAAAATTAGCCAGCGCTGGTGGTGGGCACCTGTACTGGGAGGTGCCACCCAGC 2357
DB 2298 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2826 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2768
QY 2358 TACTGGGAGGCTGAGTCAAGGAGATCGCTTGAACCTGGGAGCGGAGTTGCGGTGAGC 2417
DB 2358 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2767 TACTTCGAGGCTGAGGCGAGGAGATGGCTTGAACCCCGGAGCGGAGCTTGCAGTGAGC 2708
QY 2418 TGAGATGGTGCACATGCACCTCCAGCTGGCGAAGAGAGGACTCTGTCTCCAAAAAAG 2477
DB 2418 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2707 CGAGATCATGCCCATACACTCCAGCGCTGGGTGACAGATGAGACTCTGTCAAAAAAAGA 2648
QY 2478 AGAAGAGGA 2486
DB 2647 ACGAAAGTA 2639
```

Search completed: October 31, 1999, 01:41:38
Job time: 3367 sec

This Page Blank (uspto)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ BLACK BORDERS

☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

☐ FADED TEXT OR DRAWING

☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING

☐ SKEWED/SLANTED IMAGES

☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS

☐ GRAY SCALE DOCUMENTS

☐ LINES OR MARKS ON ORIGINAL DOCUMENT

☒ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

This Page Blank (uspto)